

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ ІМ. ІГОРЯ СІКОРСЬКОГО”
Факультет електроніки
Кафедра електронної інженерії

"На правах рукопису"

УДК 61:681.5

«До захисту допущено»
Завідувач кафедри

_____ В.І. Тимофєєв
“ ” _____ 20 р.

Магістерська дисертація

зі спеціальності 153 мікро- та наносистемна техніка

на тему «Комбіноване машинне навчання процедур розпізнавання медико-біологічних сигналів»

Виконав: студент ІІ курсу, групи ДМ-92мп
Лагутін Віталій Васильович _____
(прізвище, ім'я, по батькові) (підпис)

Науковий керівник доц. каф. ЕІ, доц., к.т.н., Шуляк О. П. _____
(посада, вчене звання, науковий ступінь, прізвище та ініціали) (підпис)

Рецензент проф. каф. АМЕС, проф., д.т.н., Продеус А. М. _____
(посада, вчене звання, науковий ступінь, прізвище та ініціали) (підпис)

Засвідчую, що у цій магістерській дисертації немає
запозичень з праць інших авторів без відповідних
посилань.

Студент _____
(підпис)

Київ - 2020 року

Форма № Н-9.01

Національний технічний університет України
“Київський політехнічний інститут імені Ігоря Сікорського”

Факультет електроніки
 Кафедра електронної інженерії
 Рівень вищої освіти – другий (магістерський)
 за освітньо-професійною програмою
 Електронні мікро- і наносистеми та технології
 Спеціальність 153 мікро- та наносистемна техніка

ЗАТВЕРДЖУЮ**Завідувач кафедри**

_____ В. І. Тимофєєв

“04” листопада 2020 р.**З А В Д А Н Н Я****НА МАГІСТЕРСЬКУ ДИСЕРТАЦІЮ СТУДЕНТУ**Лагутіну Віталію Васильовичу

(прізвище, ім'я, по батькові)

1. Тема дисертації «Комбіноване машинне навчання процедур розпізнавання медико-біологічних сигналів»

Науковий керівник Шуляк Олександр Петрович, к.т.н., доц. каф. ЕІ,

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від 05.11.2020 № 3241-с.

2. Строк подання студентом дисертації 07 грудня 2020 р.

3. Вихідні дані до роботи: 1. Прийняття рішень про тип сигналу – за їх формою зі скалярним добутком сигналів і еталонів як критерію їх порівняння. 2. Ознаки для розпізнавання сигналів – послідовності відліків характеристики їх форми. 3. Комбінування у навчанні – комбінування навчання з учителем з самонавчанням.

4. Метрики кластеризації – Евклідова відстань та показник α_z . 5. Сигнали тестового прикладу – QRS-комплекси N, A і V типів електрокардіограми пацієнта в запису з бази даних в Інтернет. Обсяги навчальних і контрольних вибірок – однакові. 6. Середня кількість сигналів в кластерах – 10. 7. Показники якості розпізнавання сигналів – чутливість, специфічність. 8. Процедури і розрахунки – в середовищі MatLab.

4. Зміст дипломної роботи (перелік питань, які потрібно розробити): 1. Завдання комбінованого навчання процедур розпізнавання медико-біологічних сигналів та розпізнавальний алгоритм для перевірки результативності його комбінованого

навчання 2. Програмні процедури алгоритму розпізнавання сигналів із його комбінованим навчанням. 3. Чисельний експеримент з оцінки результативності комбінованого навчання алгоритмів розпізнавання медико-біологічних сигналів. 4. Розробка стартап-проекту.

5. 1. Блок-схеми розроблених процедур навчання, розпізнавання сигналів і статистичних досліджень. 2. Графіки досліджуваних сигналів і характеристик їх форми. 3. Таблиці, діаграми і графіки результатів досліджень.

6. Консультанти розділів дисертації

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв

7. Дата видачі завдання – 04 листопада 2020 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів виконання магістерської дисертації	Строк виконання етапів роботи	Примітка
1	Збір і аналіз літератури	04 – 10.11	
2	Розробка розділу 1	06 – 14.11	
3	Розробка розділу 2	14 – 18.11	
4	Розробка розділу 3	18 – 26.11	
5	Розробка стартап-проекту	26 – 30.11	
6	Оформлення пояснювальної записки	04.11 – 02.12	
7	Розробка презентації	02.12 – 07.12	

Студент _____
(підпис)

Лагутін В.В.
(прізвище та ініціали)

Керівник роботи _____
(підпис)

Шуляк О.П.
(прізвище та ініціали)

РЕФЕРАТ

Магістерська дисертація: на 114 с., 4 розд., 32 табл., 30 рис., 33 джерела.

МЕДИКО-БІОЛОГІЧНІ СИГНАЛИ, АЛГОРИТМИ РОЗПІЗНАВАННЯ,
КОМБІНОВАНЕ МАШИННЕ НАВЧАННЯ, СПОСІБ І ПРОГРАМНЕ
ЗАБЕЗПЕЧЕННЯ

Об'єктом досліджень і розробок в роботі є процедури і програмне забезпечення машинного навчання алгоритмів розпізнавання медико-біологічних сигналів за їх формою в системах діагностики пацієнтів. Предмет досліджень та розробок – розробка і програмна реалізація варіанта машинного навчання розпізнавального алгоритму, у якому навчання з учителем комбінується із самонавчанням шляхом кластеризації навчальних вибірок сигналів з традиційною медичною класифікацією.

Метою роботи є створення додаткового математичного забезпечення і програмного інструментарію для виявлення, оцінки і реалізації резервів у підвищенні показників якості розпізнавання медико-біологічних сигналів за їх формою в системах діагностики пацієнтів за рахунок більш детального використання апріорної інформації про особливості розподілу таких сигналів в навчальних вибірках відомої досліднику медичної класифікації.

В першому розділі роботи розглядається комплекс часткових завдань для її виконання, конкретизуються типи сигналів і розпізнавальний алгоритм для його комбінованого навчання, пропонується сам спосіб такого навчання, програмна реалізація якого розкривається в розділі 2. Розроблені програмні процедури наведені в додатках.

Розділ 3 розкриває експериментальну перевірку розроблених алгоритмів на тестовому прикладі. У заключному розділі розглядається стартап-проект – варіант можливої комерціалізації та введення на ринок продукту, утвореного на основі виконаних досліджень.

ABSTRACT

Master's dissertation: 114 pages, 4 chapters, 32 tables, 30 figures, 33 sources.

BIOMEDICAL SIGNALS, RECOGNITION ALGORITHMS, COMBINED MACHINE LEARNING, METHOD AND SOFTWARE

The strategic line of developing is improvement of algorithms and the software of machine learning of procedures of recognition of medico-biological signals on their form in systems of diagnostics of patients.

The aim of the work is creating additional mathematical software and software tools for identifying, evaluating and implementing reserves to improve the quality of recognition of human biomedical signals by their form in patient diagnostic systems through more detailed use of a priori information about such signals in educational samples known to medical researchers. classification.

The relevance of the topic of this work is determined by the possible significant impact of the quality of recognition of medical and biological signals on the quality of diagnosis and treatment of patients in general, so the development of additional software to disclose existing and unused reserves.

The object of research and development in the work are procedures and software for machine learning algorithms for recognition of biomedical signals by their form in patient diagnostic systems.

The subject of research are the development and software implementation of a variant of machine learning recognition algorithm for biomedical signals of a given classification according to their form with combining his learning with a teacher with clustering of educational samples, which is inherent in self-learning. increasing the sensitivity and specificity of their recognition, as well as, at the same time, working out the developed procedures and confirming the effectiveness of their use on a test example of processing a real record of the patient's electrocardiogram from the Internet database.

The first section specifies the types of signals (QRS-complexes N, A and V types) and the recognition algorithm for its combined learning, the method of such learning is

offered, a set of additional tasks is considered, such as fragmentation of continuous recording of the patient's electrocardiogram, synchronization of QRS-complexes in a single observation window, pre-processing of signals before recognizing and refining the algorithm for deciding on the type of input signal, which is based on the algorithm for deciding on the type of signal according to its form for the system that learns from the teacher. In addition, based on the results of the analysis of clustering methods, a convenient clustering method was chosen for its application in the blended learning algorithm. The second section of the work presents the software implementation of the proposed method of combined learning and accompanying algorithms for pre-processing signals, preparation of learning and control samples and a refined decision-making procedure. Software implementation of algorithms is carried out in the MatLab environment, listings of the developed programs are given in appendices of work. The third section of the work reveals the content, procedure, applied criteria and algorithm of test verification of the teaching method and developed software procedures. The test of the learning algorithm and its accompanying programs takes place on a real half-hour recording of the patient's ECG signal, which is taken continuously from the Internet database. The results of the test for clarity are presented in the format of tables and charts. The fourth section "Development of a startup project" reveals one of the possible areas of formation of an independent product based on research and includes an assessment of the main characteristics and benefits of the product, the market situation. Potential competitors and consumers of the product are also reviewed, to which is added the general marketing program of the product.

In conclusions reveals the content of the main results of its implementation and the main provisions (conclusions) that follow from them.

In general, the results of the development are proposed to be considered as additional tools for identifying in the analysis of patient data the presence, evaluation and implementation of reserves in improving the validity of recognition of biomedical signals used to diagnose these patients.

ЗМІСТ

ВСТУП.....	9
1 ЗАВДАННЯ КОМБІНОВАНОГО НАВЧАННЯ ПРОЦЕДУР РОЗПІЗНАВАННЯ МЕДИКО-БІОЛОГІЧНИХ СИГНАЛІВ ТА РОЗПІЗНАВАЛЬНИЙ АЛГОРИТМ ДЛЯ ПЕРЕВІРКИ РЕЗУЛЬТАТИВНОСТІ ЙОГО КОМБІНОВАНОГО НАВЧАННЯ.....	15
1.1 Завдання комбінованого навчання процедур розпізнавання медико-біологічних сигналів та загальна схема його реалізації і перевірки результативності.....	15
1.2 Характеристики сигналів, що розпізнаються і алгоритм розпізнавання сигналів для його комбінованого навчання	20
1.3 Порядок формування навчальних і контрольних вибірок для навчання алгоритмів розпізнавання сигналів і оцінки його результативності.	26
1.4 Ознаки сигналів для їх розпізнавання і зміст попередньої обробки даних для їх розрахунку	30
1.5 Принцип і порядок комбінованого навчання розпізнавальних алгоритмів	31
1.6 Висновки до розділу 1	43
2 ПРОГРАМНІ ПРОЦЕДУРИ АЛГОРИТМУ РОЗПІЗНАВАННЯ СИГНАЛІВ ІЗ ЙОГО КОМБІНОВАНИМ НАВЧАННЯМ	48
2.1 Процедура формування навчальних і контрольних вибірок сигналів	48
2.2 Процедури попередньої обробки вхідних сигналів.....	50
2.3 Процедури кластеризації навчальних вибірок та формування еталонів.	50
2.4 Процедура прийняття рішення про тип вхідного сигналу	55
2.5 Висновки до розділу 2	58

3 ЧИСЕЛЬНИЙ ЕКСПЕРИМЕНТ З ОЦІНКИ РЕЗУЛЬТАТИВНОСТІ КОМБІНОВАНОГО НАВЧАННЯ АЛГОРИТМІВ РОЗПІЗНАВАННЯ МЕДИКО-БІОЛОГІЧНИХ СИГНАЛІВ	60
3.1 Мета і завдання чисельного експерименту	60
3.2 Критерії статистичної оцінки результативності навчання алгоритму розпізнавання сигналів	62
3.3 Дані для проведення чисельного експерименту	63
3.4 Програмні процедури середовища MATLAB для проведення чисельного експерименту	66
3.5 Результати чисельного експерименту та їх аналіз.....	67
3.6 Висновки до розділу 3	80
РОЗРОБКА СТАРТАП-ПРОЕКТУ.....	82
4.1 Опис ідеї проекту	82
4.2 Технологічний аудит ідеї проекту.....	84
4.3 Аналіз ринкових можливостей запуску стартап-проекту.....	85
4.4 Аналіз ринкових можливостей запуску стартап-проекту.....	91
4.5 Розроблення маркетингової програми стартап-проекту.....	94
4.6 Висновки до розділу «Розробка стартап-проекту».....	97
ВИСНОВКИ.....	98
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ.....	101
Додаток А	105
Додаток Б.....	109
Додаток В	111

ВСТУП

Одним із важливих напрямів розвитку медичних діагностичних систем є удосконалення алгоритмів розпізнавання типів медико-біологічних сигналів згідно з різними їх класифікаціями, які були зроблені відповідно до станів організму людини, окремих його органів і систем, відповідно до захворювань різних видів, стадій їх розвитку, а також змісту здійснюваних медичних втручань та етапів одужання пацієнтів [1-5].

Значні зусилля в такому напрямку прикладаються до удосконалення знань про ці сигнали [1-5], для чого розвинутими розпізнавальними системами здійснюється систематичне накопичення апріорної інформації про вигляд характерних сигналів різних типів у їх використовуваних класифікаціях, формуються системи різноманітних характеристик, параметрів і описів у вигляді комплексів ознак для їх розпізнавання [1-5]. Суттєве значення при цьому має удосконалення методів більш ретельного і повного відшукування в апріорних даних корисної інформації про типові сигнали медично інтерпретованих класифікацій [1-5], завдяки чому все більш повно розкриваються і реалізуються наявні резерви підвищення рівня правильності їх розпізнавання і діагностики і лікування пацієнтів в цілому.

Досягнення та результати усіх цих удосконалень впроваджуються в системах обробки даних на етапі навчання алгоритмів розпізнавання сигналів [1, 2, 4, 5], на якому саме й проводиться необхідна обробка потрібної апріорної інформації про конкретні типи сигналів, що розпізнаються, здійснюється налаштування розпізнавальних алгоритмів на її використання, враховуються формати опису цих сигналів, оцінюються чутливість, специфічність, інші досягнуті показники якості розпізнавання сигналів, формуються рекомендації для використання модифікованих алгоритмів [1-5]. Тому завдання удосконалення програмного забезпечення для машинного навчання таких розпізнавальних алгоритмів в системах діагностики пацієнтів, у тому числі, за рахунок комбінування методів їх навчання для розкриття резервів підвищення показників вірності розпізнавання

медико-біологічних сигналів людини, зберігає свою актуальність і тема роботи актуальна.

Дослідження і розробки в цій роботі щодо комбінування навчання проводились стосовно широко розповсюджених алгоритмів визначення типу сигналів певної класифікації за їх формою [1-6]. Для них широко розповсюджене визначення для кожного класу по навчальних вибірках сигналів характерної форми (еталонів) [1, 5, 6, 8-14]. При цьому частіш за все для кожного класу використовується один еталон [1, 5, 6, 8-14], який узагальнює наявні апріорні дані. Рішення про тип вхідного сигналу приймається шляхом порівняння його форми з формою еталонів. Вважається, що сигнал надійшов з того класу, до еталону якого він є найближчим за своєю формою [1, 6, 9-14].

Еталоном в кожному класі вважається середній вигляд сигналів навчальної вибірки цього класу [1, 6, 9-14]. І це має місце в умовах, коли, відповідно до результатів численних досліджень, реальні розподіли по вірогідності зображуючих медико-біологічні сигнали точок у відповідних просторах і гістограмах мають різноманітну та, як правило, складну форму [1-6]. Використання одиничних еталонів по класах сигналів, усереднення екземплярів сигналів навчальних вибірок під час їх розрахунку унеможливають в такому разі детальне врахування особливостей цих розподілів з їх скупченнями і розрідженнями під час визначення типу вхідних сигналів. Тому в роботі вважається, що більш детальне врахування особливостей реальних розподілів сигналів може сприяти підвищенню рівня правильності їх розпізнавання.

Особливість розпізнавання сигналів з використанням в кожному їх класі єдиного еталону полягає в тому, що під час його формування шляхом усереднення наявних екземплярів сигналу навчальної вибірки на етапі навчання розпізнавальних алгоритмів, втрачається можливість враховувати для забезпечення високого рівня правильності розпізнавання сигналів особливостей реального розподілу їх значень по ймовірності з наявними скупченнями, розрідженнями зображуючих точок, можливою полімодівістю і асиметрією, які

характерні для багатьох медико-біологічних сигналів і характеристик біологічних матеріалів [1-6, 9, 11, 13, 14].

Завданням цієї роботи було відновити зазначену можливість, розробити і програмно реалізувати такі процедури навчання розпізнавального алгоритму, в тому числі базового, взятого для прикладу для розкриття змісту потрібних дій, який міг би забезпечувати отримання з апіорних даних навчальних вибірок інформації про реальні особливості ймовірнісного розподілу сигналів різних класів та подавав їх у зручному форматі для використання алгоритмом розпізнавання під час його роботи після навчання.

В цілому, все це було спрямовано на створення програмного інструментарію виявлення, оцінки і реалізації резервів підвищення вірності розпізнавання типів сигналів за рахунок більш детального використання розпізнавальними алгоритмами апіорної інформації з наявних навчальних вибірок сигналів кожного їх класу.

Для вирішення завдання розробки навчального алгоритму з метою детального врахування розподілів сигналів було застосовано такий спосіб. Спочатку кожна навчальна множина кожного типу сигналів розбивається на менші підмножини екземплярів за ознакою близькості зображуючих їх точок. Дислокації точок таких множин відповідають, таким чином, локаціям скупчень в їх реальних розподілах та вбирають в себе і характеризують в такій формі важливі для розпізнавання сигналів особливості їх розподілів в кожному класі. Потім для кожної такої групи знову ж такі усередненням близьких екземплярів розраховуються сімейства еталонів. І, на сам кінець, усі отримані таким чином групи еталонів усіх класів використовуваної класифікації передаються для використання від навчального алгоритму до алгоритму розпізнавального. В цьому полягає зміст способу і алгоритму навчання, що пропонується в роботі.

Розбивка множин сигналів навчальних вибірок на підмножини близьких екземплярів в місцях їх скупчення в наявних дислокаціях зображуючих точок є типовим завданням кластеризації множин точок, що зображують сигнали [2, 1, 3, 5, 4, 8]. Для вирішення такого завдання вже розроблені різноманітні математичні і

програмні процедури, які використовуються на практиці [2, 1, 3, 5, 6, 4] і знайшли використання в цій роботі. При цьому слід зауважити, що з іншого боку, з боку класифікацій методів навчання розпізнавальних алгоритмів, серед основних з котрих перелічуються навчання з учителем і самонавчання [2, 1, 3, 5, 6, 4], кластеризація множин сигналів, їх розбивка на підмножини близьких екземплярів, є типовою дією процесу самонавчання [15, 1, 3, 5]. Тому формування навчальних вибірок сигналів учителем згідно з медично інтерпретованою їх класифікацією для алгоритму, який навчається, в комбінації з властивою самонавчанню кластеризацією цих вибірок для врахування в такій формі особливостей розподілів сигналів в класах з метою підвищення вірності їх розпізнавання розглядається в роботі як комбінована процедура навчання розпізнаванню сигналів з ознаками як навчання з учителем, так і самонавчання.

Зміст задання на комбіноване навчання розпізнавального алгоритму, принцип та відповідні процедури його навчання, які покладені в основу навчального алгоритму та використовуються в подальшому для їх програмної реалізації та дослідження результативності навчання, розглядаються в розділі 1 роботи. Визначено також, яким є алгоритм розпізнавання сигналів, що навчається з учителем і використовується як початковий, базовий для ілюстрації його комбінованого навчання, а саме, для розкриття змісту виконуваних дій процедури цього навчання та отримання порівняльних оцінок результативності зазначеного комбінування в проведенні навчання.

Конкретизовано також правило прийняття рішень щодо розпізнавання сигналів, яке базується на порівнянні вхідного сигналу з еталонами та передбачає вибір в якості рішення такого класу сигналів, який дає максимум цієї подібності за їх скалярним добутком [6, 9, 11-14] після перетворення сигналів до послідовностей відліків характеристики їх форми [9, 11-14]. Під такою характеристикою в цілях цієї роботи розуміється таж сама послідовність відліків сигналу в початковому його вигляді, з яких вилучено постійну складову та які піддані додатковому масштабуванню [11, 13, 14].

Це зроблено, по-перше, для того, щоб полегшити умови розпізнавання сигналів та, по-друге, для того, щоб використання скалярного добутку сигналу та еталону давало коректний результат порівняння їх форми [6, 11, 13, 14].

Визначається, крім того, кількість класів сигналів, що розпізнаються, надаються їх тип і характеристика (N, A і V типи QRS-комплексів електрокардіограми пацієнта [8, 9, 11-14]), формати їх подання, а також особливості спостереження цих сигналів і формування навчальних вибірок. Вибір цих сигналів для ілюстрації комбінованого навчання базового розпізнавального алгоритму продиктований виключно наявністю достатньої кількості гарантовано класифікованих екземплярів сигналів в базі даних в Internet [16, 17] для формування навчальних вибірок і здійснення навчання.

Другий розділ роботи присвячений програмній реалізації комплексу розроблених процедур, які по'язані з темою комбінованого машинного навчання процедур розпізнавання медико-біологічних сигналів. Вирішення цього завдання здійснюється в середовищі MatLab [18, 19].

Пропонується структура і програмна реалізація таких обчислюваних схем: процедури попередньої обробки даних для формування учителем навчальних вибірок сигналів, що розпізнаються, та їх перетворення до послідовностей відліків характеристик форми; процедури базового алгоритму розпізнавання сигналів, що навчається з учителем, який взято для ілюстрації, та декілька іншої його модифікація у зв'язку з переходом до його комбіноване навчання з багато еталонним описом навчальних вибірок; процедури навчання алгоритму розпізнавання сигналів, який охоплює етап його навчання з визначенням кластерів у складі навчальних вибірок, формуванням локальних еталонів кожного класу сигналів та передавання цих еталонів базовому алгоритму, який навчається.

В третьому розділі роботи надається опис перевірки та відпрацювання розроблених програмних процедур, розглядаються умови і результати щодо порівняльної оцінки результативності комбінованого навчання базового алгоритму

навчання за допомогою розроблених процедур його комбінованого навчання і порівнянні з його навчанням за схемою навчання з учителем.

В розділі розкриваються цілі, завдання, умови і результати чисельних досліджень щодо комбінованого навчання згаданого навчального алгоритму. Наводяться результати цих досліджень, проводиться їх аналіз з відповідними висновками.

Четвертий розділ роботи – стартап-проект [20-23].

В цілому, надано алгоритми і програмні процедури навчання для алгоритмів розпізнавання сигналів за їх формою, які легко конкретизувати і переналагодити на іншу кількість типів сигналів в їх класифікації, на їх конкретний вигляд і формат представлення, щоб в подальшому оцінити можливий виграш у вірності розпізнавання типів сигналів та реалізовувати цей виграш в подальшому в системах обробки медико-біологічних сигналів пацієнтів для підвищення якості їх діагностики.

Результати досліджень і розробок за тематикою, яка пов'язана з темою дисертації, опубліковано за участі автора в ряді наукових робіт [8, 10] та викладено в доповідях на науково-технічних конференціях [9, 24, 25, 26].

1 ЗАВДАННЯ КОМБІНОВАНОГО НАВЧАННЯ ПРОЦЕДУР РОЗПІЗНАВАННЯ МЕДИКО-БІОЛОГІЧНИХ СИГНАЛІВ ТА РОЗПІЗНАВАЛЬНИЙ АЛГОРИТМ ДЛЯ ПЕРЕВІРКИ РЕЗУЛЬТАТИВНОСТІ ЙОГО КОМБІНОВАНОГО НАВЧАННЯ

1.1 Завдання комбінованого навчання процедур розпізнавання медико-біологічних сигналів та загальна схема його реалізації і перевірки результативності

Одним із важливих напрямів підвищення вірності розпізнавання медико-біологічних сигналів, від чого суттєво залежить якість діагностики пацієнтів в цілому, є більш ретельний аналіз апіорних даних на етапі навчання розпізнавальних алгоритмів, отримання і використання більш детальної інформації про ці сигнали з навчальних вибірок [1-5]. Завдяки цьому навіть досить прості алгоритми прийняття рішень про типи сигналів, що розпізнаються, підвищують чутливість, специфічність, загальну валідність в їх розпізнаванні [1-5, 9, 13, 14].

Валідність рішень розпізнавального алгоритму тут розуміється як кількісний показник відповідності цих рішень до дійсності, тобто такий показник, що одночасно враховує рівень правильності рішень стосовно типу вхідних сигналів та, у той же час, частку правильних рішень серед усіх, які прийняті цим алгоритмом. Відокремлено ці два показники розуміються як чутливість і специфічність розпізнавання сигналів за допомогою певного розпізнавального алгоритму, що розглядається.

Ці показники перевіряються і підтверджуються в практиці обробки даних зазвичай статистично шляхом безпосереднього випробування розпізнавальних процедур на контрольних вибірках [1, 3-5] після їх модифікації, і це потребує проведення додаткової роботи щодо підтвердження наявності позитивного результату в підвищенні якості роботи модифікованого розпізнавального алгоритму, тобто і в підвищенні якості його навчання, а саме, в забезпеченні отримання додаткової корисної для розпізнавання сигналів інформації з наявних апіорних даних навчальних вибірок.

Тому завдання комбінованого навчання процедур розпізнавання медико-біологічних сигналів в роботі розглядається як комплексне завдання, яке передбачає не тільки саме розробку алгоритму навчання і його програмну реалізацію, але й статистичне підтвердження результативності використаного способу комбінування навчання з учителем [1, 3-5] з самонавчанням [1, 3-5] шляхом кластеризації навчальних вибірок [8, 27-31] в класах сигналів, що розпізнаються, і подальшого використання багато еталонного опису таких вибірок для більш повного врахування реальних ймовірнісних розподілів сигналів в кожному їх класі.

В зв'язку з цим у складі цього завдання було потрібно передбачити також вибір зручного типу самого алгоритму розпізнавання сигналів, який навчається зазначеним чином за допомогою розроблених процедур, вибір типу медичних сигналів і формату їх представлення для демонстрації змісту їх обробки під час пробного навчання, конкретизацію використовуваної класифікації цих сигналів, а також складу ознак кожного з них для проведення розпізнавання, формування навчальних і контрольних вибірок та, за необхідності, проведення потрібної попередньої обробки сигналів, організацію статистичної порівняльної оцінки якості їх розпізнавання початковим і модифікованим в роботі алгоритмами і вирішити інші супутні питання.

Слід також зазначити, що в зв'язку з переходом до багато еталонного опису навчальних вибірок потребує також деякого доопрацювання процедура прийняття рішень в розпізнавальному алгоритмі, який використовується в дослідженнях якості комбінованого навчання і який доцільно достатньо простим, тобто таким, що для опису сигналів кожного класу використовує лише одиничний характерний вигляд сигналу, так званий еталон [1, 5, 6, 8-14], а рішення щодо типу сигналів в ньому приймається шляхом порівняння вхідного сигналу з еталонами [1, 5, 6, 8-14] за критерієм їх близькості на користь найближчого еталону [1, 5, 6, 8-14]. Усі перелічені для виконання роботи математичні і логічні процедури мають бути розроблені і представлені у зручній для програмної реалізації формі в середовищі MATLAB.

Таким чином, в цілому завдання комбінованого навчання процедур розпізнавання медико-біологічних сигналів в цій роботі розглядається як сукупність таких взаємопов'язаних часткових завдань (рис. 1.1).

Визначення типу алгоритму, що навчається з учителем, для пробного його комбінованого навчання здійснюється з міркувань простоти і забезпечення наочності його навчання. Алгоритм використовується для ілюстрації змісту його комбінованого навчання та задля проведення порівняльних досліджень

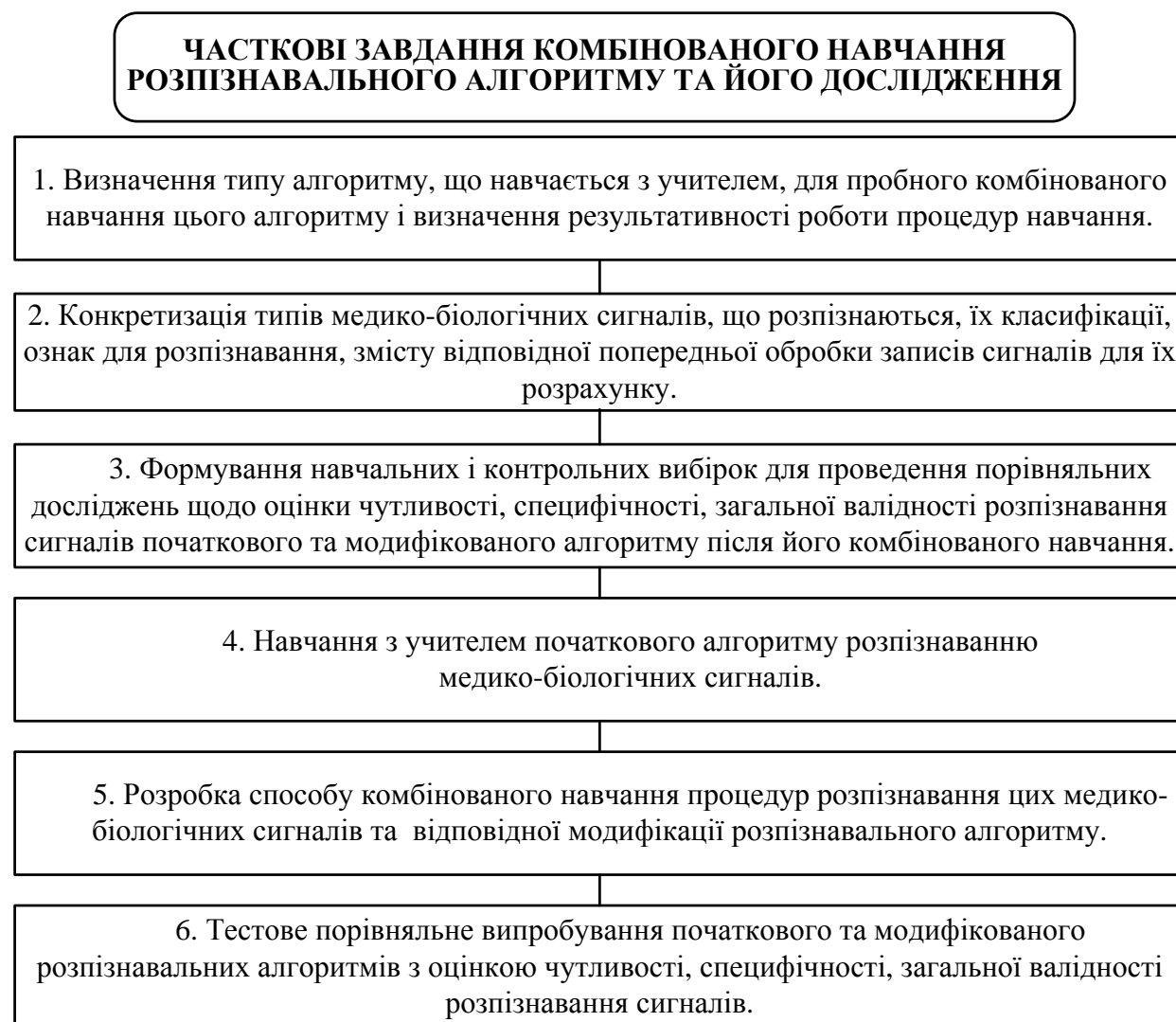


Рисунок 1.1 – Часткові завдання комбінованого навчання розпізнавального алгоритму та його дослідження

результативності навчання в початковому та модифікованому вигляді.

Алгоритм потребує налаштування на етапі навчання на конкретний вид медико-біологічних сигналів, що розпізнаються, їх класифікацію з певною кількістю класів, систему ознак для їх розпізнавання та формат їх представлення в процедурах обробки даних. Зміст цього алгоритму та згадані характеристики сигналів, що розпізнаються, розглядаються в пункті 1.2 цієї роботи. Порядок визначення результативності роботи запропонованих процедур комбінованого навчання, а також отримані з цього приводу результати наводяться в розділі 3.

У якості ознак сигналів для їх розпізнавання використовуються послідовності відліків характеристики їх форми [9, 11-14]. Зміст попередньої обробки записів сигналів для формування навчальних і контрольних вибірок для навчання алгоритмів розпізнавання сигналів і оцінки його результативності розкривається в пункті 1.3, а склад ознак сигналів для їх розпізнавання і зміст попередньої обробки даних для їх розрахунку – в пункті 1.4. Обробка записів сигналів з бази даних для формування навчальних і контрольних вибірок полягає в фрагментації цих записів, приведення їх до єдиного вікна спостереження і в сортуванні сигналів по класам згідно з використовуваною класифікацією. Початковою формою запису цих сигналів є послідовності їх відліків на рівномірній сітці часу. Система ознак сигналів для їх розпізнавання – це також послідовність відліків, але характеристики їх форми. Зміст поняття характеристики форми сигналу також розкривається в пункті 1.4.

Порядок оцінки чутливості, специфічності, загальної валідності розпізнавання сигналів досліджуваними алгоритмами обговорюється в розділі 3 роботи.

Принцип і порядок комбінованого навчання розпізнавальних алгоритмів наводиться в пункті 1.5. Зміст такого навчання полягає в кластеризації навчальних вибірок сигналів різних класів, що розглядаються, і визначенні еталонів в кожному з цих класів. Кожна навчальна вибірка в такому разі характеризується групою еталонів, склад якої залежить від особливостей імовірнісного розподілу її сигналів.

У разі навчання алгоритмів в початковому вигляді для кожного класу навчальної вибірки визначається одиничний еталон.

Тестові порівняльні випробування початкового та модифікованого розпізнавальних алгоритмів з оцінкою чутливості, специфічності, загальної валідності розпізнавання сигналів розглядаються в розділі 3 роботи..

Таким чином, визначено склад та деякі особливості основних часткових завдань комплексного завдання комбінованого навчання процедур розпізнавання медико-біологічних сигналів в цій роботі – основних процедур загальної схеми проведення цього навчання і перевірки його результативності. Зазначено розподіл виконання цих часткових завдань по пунктах в роботі. Загальна схема реалізації і перевірки результативності комбінованого навчання розпізнавального алгоритму та його дослідження представлена на рисунку 1.2.

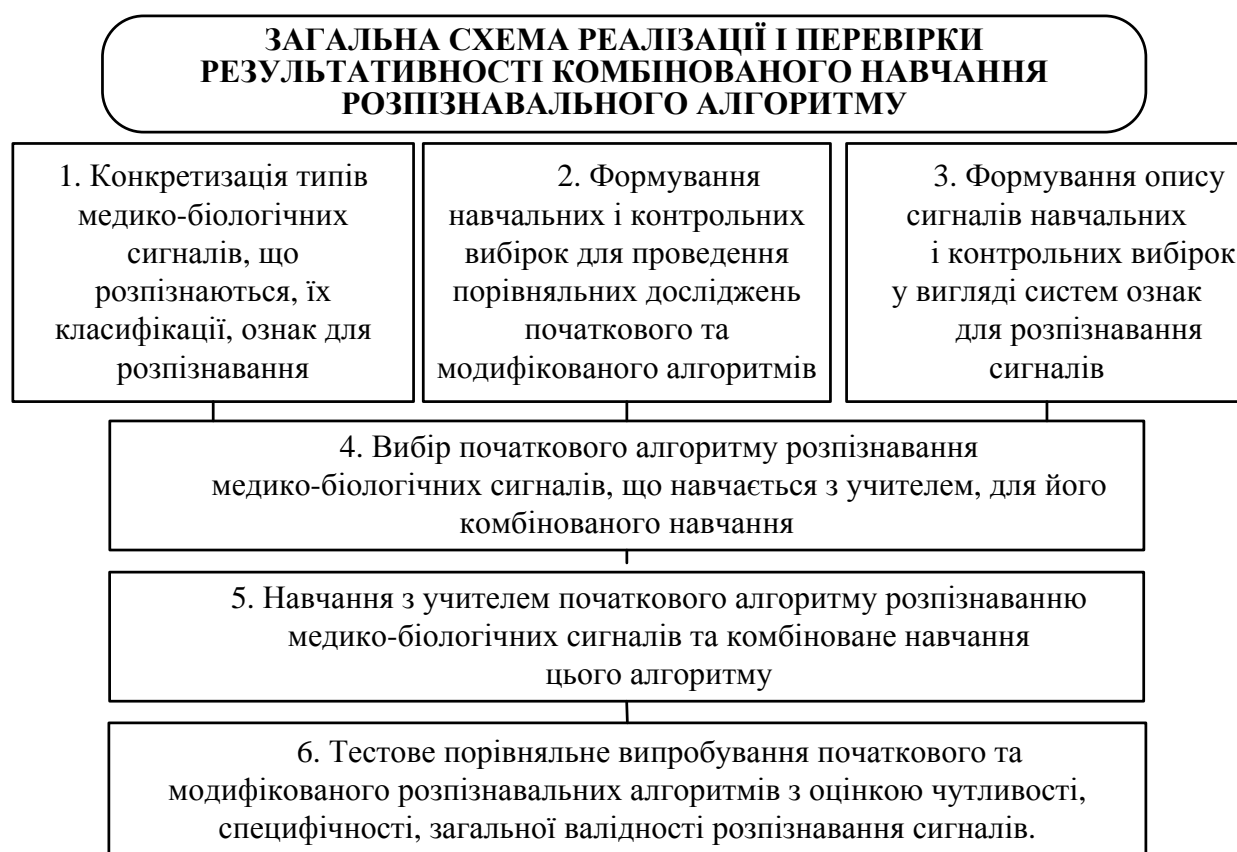


Рисунок 1.2 – Загальна схема реалізації і перевірки результативності комбінованого навчання розпізнавального алгоритму

За цією схемою спочатку формуються сигнали для проведення досліджень, потім проводиться навчання початкового та модифікованого алгоритмів і далі

здійснюється тестова перевірка результативності застосування комбінованого способу у навчанні розпізнавального алгоритму.

Таким чином, в цілому:

- визначено склад основних часткових завдань комплексного завдання комбінованого навчання процедур розпізнавання сигналів;
- розкрито загальну схему проведення розробок і досліджень в роботі.

Розпізнавальний алгоритм, що навчається з учителем і проходить комбіноване навчання, а також характеристика сигналів і ознак для їх розпізнавання розглядаються в наступному підрозділі цієї роботи.

1.2 Характеристики сигналів, що розпізнаються і алгоритм розпізнавання сигналів для його комбінованого навчання

Розкриємо загальні міркування щодо вибору типу алгоритму для його комбінованого навчання, а також загальну інформацію про сигнали, які було взято як приклад для розкриття змісту такого навчання та дослідження його якості.

При цьому важливо мати на увазі, що розпізнавальний алгоритм в цій роботі розглядається лише як об'єкт для навчання розробленою та досліджуваною навчальною процедурою, що він використовується лише для наочного розкриття змісту комбінованого навчання на навчальних вибірках та для забезпечення оцінки результативності такого навчання статистичним шляхом на контрольних вибірках з розрахунком чутливості, специфічності, загальної валідності розпізнавання сигналів розпізнавальним алгоритмом, який пройшов комбіноване навчання.

Під час вибору розпізнавального алгоритму для досліджень комбінованого навчання врахуємо, насамперед, те, що цей алгоритм має бути таким, що він навчається з учителем. Це впливає зі змісту комбінування навчання, яке розглядається в цій роботі.

У той же час, має бути враховано, що розпізнавальні процедури, навчальний алгоритм і характеристики сигналів мають бути узгоджені між собою.

Візьмемо також до уваги, що конкретна реалізація подібного розпізнавального алгоритму для дослідження алгоритму його комбінованого навчання, в тому числі, програмна, залежить від типу медико-біологічних сигналів, кількості класів спостережуваних процесів, що розпізнаються, системи ознак для їх розпізнавання та формату їх представлення.

Стосовно сигналів, що розглядаються в роботі для проведення досліджень, та умов їх спостереження введемо такі припущення (рис. 1.3).

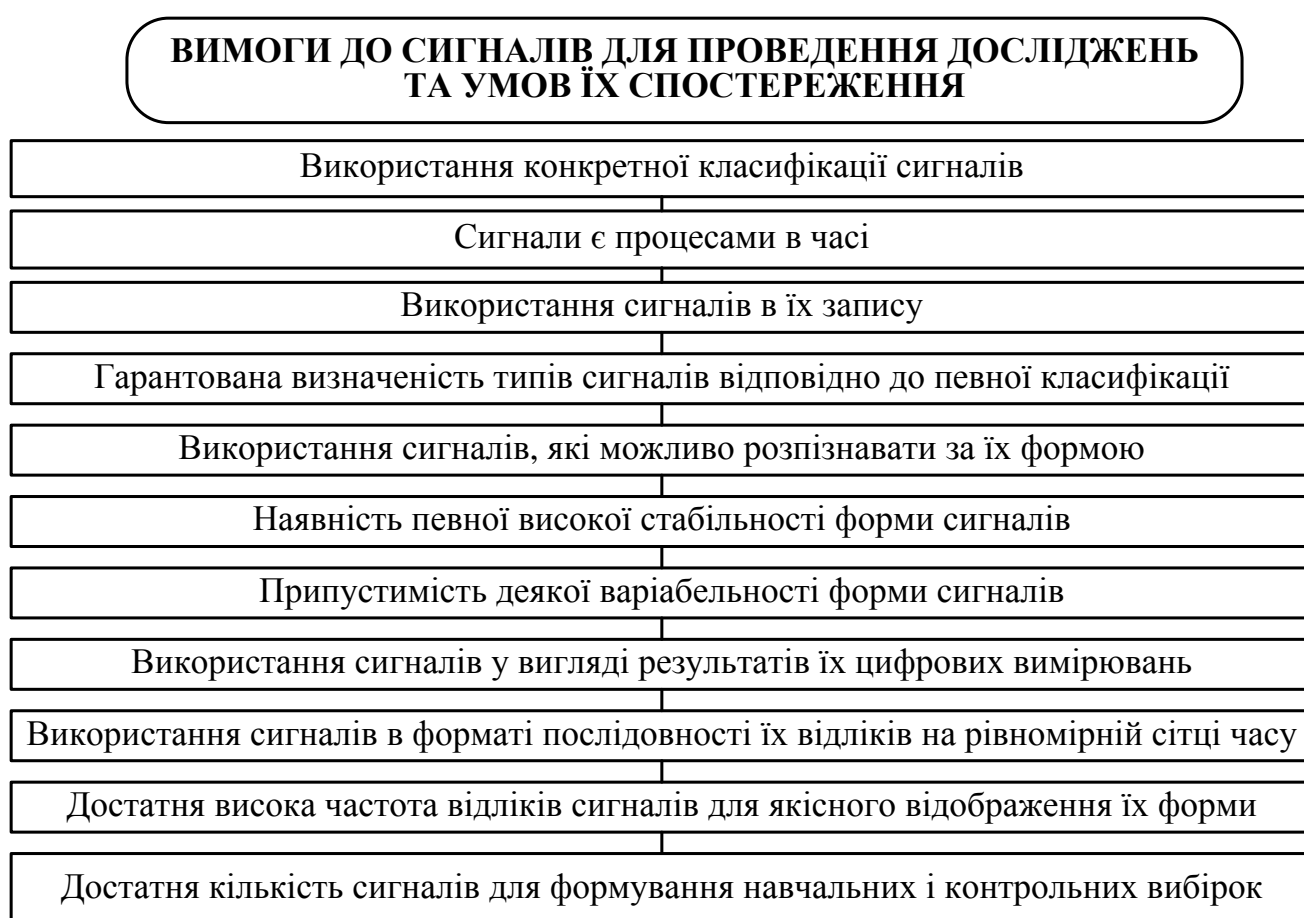


Рисунок 1.3 – Вимоги до сигналів для проведення досліджень та умов їх спостереження

Таким чином, по-перше, будемо вважати, що сигнали, є процесами, які розгортаються в часі. Спектри сигналів та інші характеристики, послідовності відліків значень яких можуть використовуватися як ознаки для їх розпізнавання,

або впорядковані списки параметрів, як наприклад, результати аналізу крові чи інших продуктів життєдіяльності організму людини в якості сигналів в системах діагностики пацієнтів в цій роботі не розглядаються.

Разом з тим, в цій роботі будемо вважати, що сигнали, які аналізуються, вимірюються цифровим методом на рівномірній сітці часу, тобто початковим виглядом кожної реалізації сигналу, який розпізнається, являє собою послідовність його числових відліків в часі на шкалі їх значень. Шкала сигналів різних класів та ж сама.

Крім того, будемо вважати, що розпізнавальний алгоритм розпізнає сигнали за їх формою, тобто для медико-біологічних сигналів для проведення розробок і досліджень має бути характерна деяка досить висока, хоча і неповна стабільність форми по класах використовуваної класифікації. Іншими словами, допускаємо певну помітну, але невелику варіабельність їх форми.

Робота орієнтована на використання записів сигналів з визначеною учителем гарантованою їх належністю відомим класам класифікації, яка передбачається, що відповідає потребам навчання розпізнавальних алгоритмів з учителем.

Важливо також, щоб для того, щоб забезпечити якісне навчання розпізнавального алгоритму та перевірку результативності проведеного комбінованого навчання і в його порівнянні з навчанням з учителем, потрібна наявність достатньої кількості екземплярів сигналів кожного класу в навчальних і контрольних вибірках.

Згідно з такими бажаними властивостями сигналів для демонстрації змісту виконуваних досліджень і розробок та розкриття отриманих результатів в роботі для конкретизації обробки даних використовуються наявні в базі даних в Інтернет півгодинні записи електрокардіограми пацієнта, в яких є достатня кількість гарантовано класифікованих фахівцями QRS-комплексів трьох типів відповідно до наведеної в цій базі медичної класифікації і використані в якості сигналів, що розпізнаються. Відліки сигналу ЕКГ тут проводились з постійним кроком на шкалі часу на достатньо високій частоті в цифровій формі.

З бази даних було взято запис з QRS-комплексами N, A і V типів, для яких на рисунку (рис. 1.4) представлена їх характерна форма в єдиному вікні

спостереження без постійної складової сигналів в нормованому по інтенсивності вигляді.

Тут під N типом QRS-комплексу розуміється сигнал електричної активності серця у разі нормального виконання ним насосної функції. Тип A відповідає передчасному скороченню серцевого м'язу раніше потрібного моменту

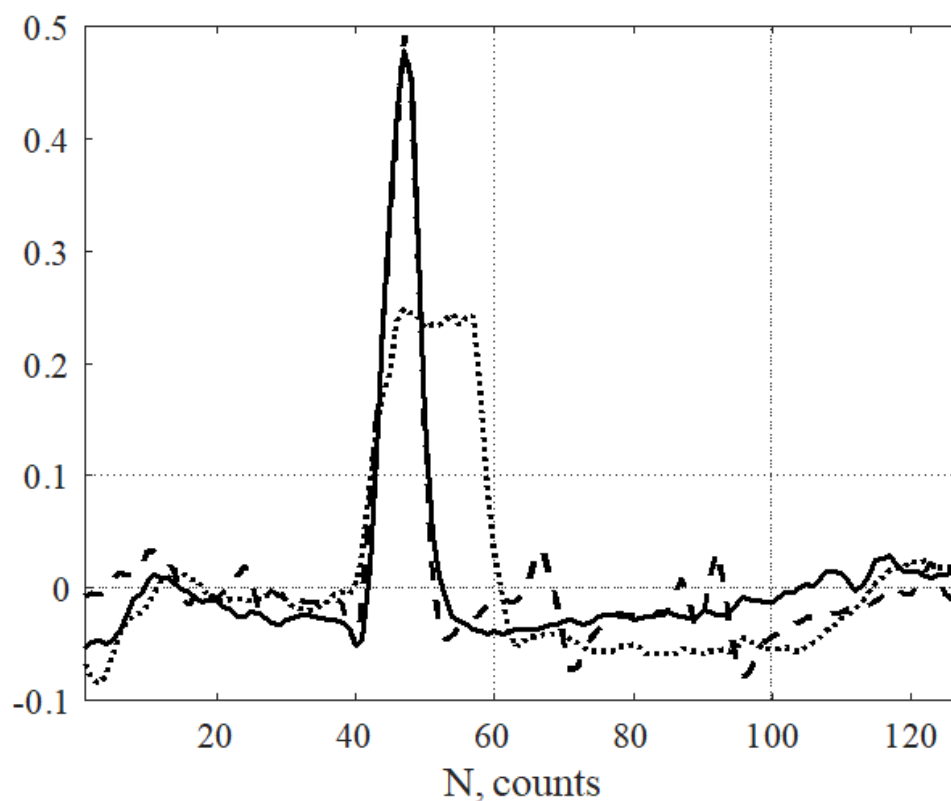


Рисунок 1.4 – Характерний вигляд QRS-комплекси N(штрих-пунктир), A (суцільна лінія) і V (пунктир-крапка) типів як сигналів для їх розпізнавання, що представлені в єдиному вікні спостереження без постійної складової сигналів в нормованому по інтенсивності вигляді

необхідного забору серцем крові для її подальшого виштовхування в коло кровообігу.

Тип V – це випадок, коли після нормального забору крові її виштовхування не трапляється, тобто потрібний R-пік в комплексі відсутній.

Використовуваний запис ЕКГ в базі даних має спеціальну розмітку про тип кожного QRS-комплексу.

Формування навчальних і контрольних вибірок для забезпечення навчання та оцінки його якості було проведено шляхом фрагментації запису ЕКГ за допомогою спеціально розробленої процедури вилученням з запису і сортування цих комплексів по вибірках з однаковою їх кількістю в кожній.

Стосовно алгоритму розпізнавання сигналів, що використовується в роботі для навчання для демонстраційних цілей, слід зауважити таке. З урахуванням перелічених особливостей сигналів, що розпізнаються, формату даних для їх представлення, а також наявних результатів досліджень, які пов'язані з тематикою цієї роботи [8-10, 24, 25, 26], та виходячи з наведених вище міркувань про доцільні властивості потрібного розпізнавального алгоритму, було вибрано для комбінованого навчання алгоритм розпізнавання сигналів, що навчається з учителем та реалізує такий спосіб розпізнавання, який структурно охарактеризовано на рисунку 1.5. Зміст кожного елементу структурного опису цього способу конкретизовано на цьому рисунку.

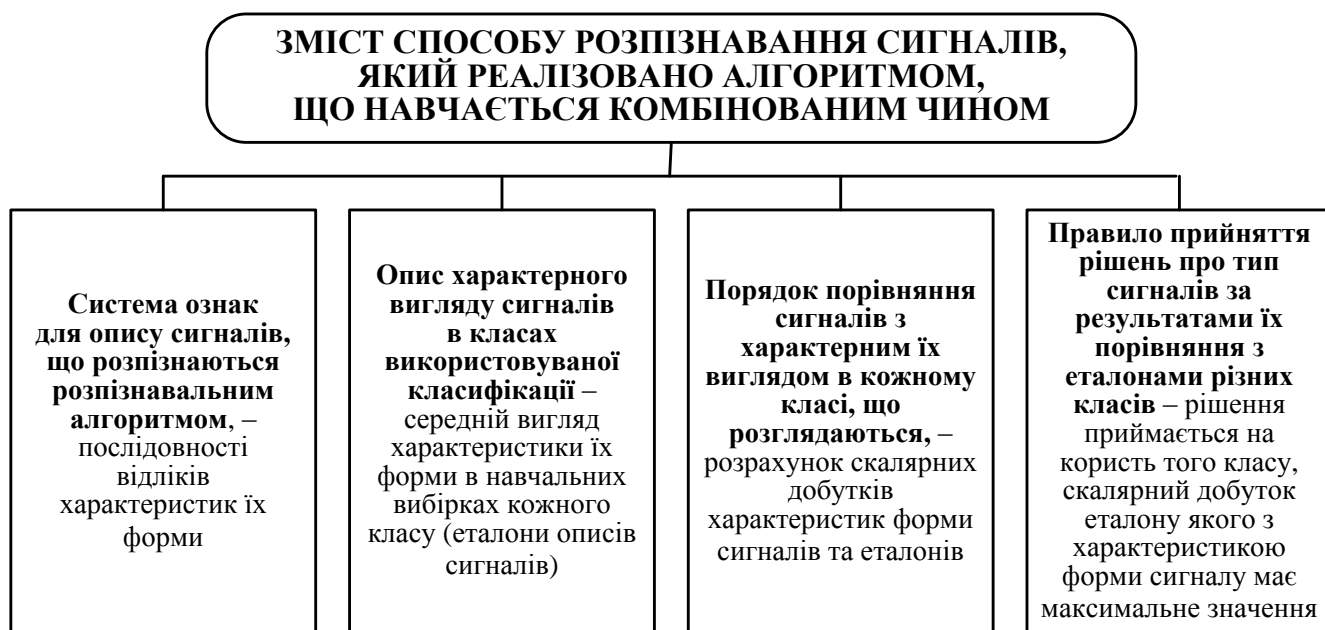


Рисунок 1.5 – Зміст способу розпізнавання сигналів, який реалізовано алгоритмом, що навчається комбінованим чином

В цілому, спосіб розпізнавання типу вхідних сигналів полягає в порівнянні їх форми з характерною формою сигналів кожного класу та в виборі того класу, еталон якого виявляє найбільшу подібність до форми сигналу, який розпізнається. Вибірки для навчання і оцінки його якості формуються в процесі попередньої обробки наявних даних з бази в Інтернет. Формування характерного вигляду сигналів в класах, що розглядаються, є завданням навчального алгоритму. Оцінка якості навчання – завдання додаткових статистичних досліджень, які розглядаються в роботі на тестовому прикладі.

Під час розробки процедур комбінованого навчання розпізнавального алгоритму доцільно було зберегти в основному його розглянутий зміст.

Таким чином, в цілому:

- розглянуто основні вимоги до сигналів для проведення досліджень по темі роботи та до умов їх спостереження і представлення їх результатів;
- визначені конкретні типи медик-біологічних сигналів для проведення розробок;
- конкретизована класифікація цих сигналів;
- виконано загальний опис розпізнавального алгоритму, що навчається з учителем, для його комбінованого навчання, зокрема, розкрито систему ознак для опису сигналів, що розпізнаються розпізнавальним алгоритмом, охарактеризовано зміст опису характерного вигляду сигналів в класах використовуваної класифікації (еталонів), конкретизовано порядок порівняння сигналів з характерним їх виглядом в кожному класі, що розглядаються, сформульовано правило прийняття рішень про тип сигналів за результатами їх порівняння з еталонів різних класів.

У зв'язку з наявними особливостями послідовності циклічно повторюваних в електрокардіограмі QRS-комплексів, зокрема, варіабельності серцевого ритму, тобто змінності частоти повторення таких сигналів, деякої варіабельності їх форми, нерегулярного коливання ізолінії, на якій розташовані такі комплекси, потрібна спеціальна процедура вилучення цих комплексів із запису ЕКГ, забезпечення єдиних умов спостереження і аналізу, сортування таких сигналів по типах. Порядок

формування навчальних і контрольних вибірок в таких умовах розкрито в наступному пункті роботи.

1.3 Порядок формування навчальних і контрольних вибірок для навчання алгоритмів розпізнавання сигналів і оцінки його результативності.

Навчальні вибірки зазначених сигналів формуються для того, щоб забезпечити розкрити змісту процедур навчання обраного алгоритму розпізнавання сигналів за схемою навчання з учителем та комбінованим способом, який розробляється в роботі, для статистичної порівняльної оцінки результативності комбінованого навчання. Крім того, навчальні вибірки потрібні також для відпрацювання та оцінки результативності подальшої роботи на помилках на етапі навчання розпізнавального алгоритму за допомогою додаткових процедур, що представлених в пункті 1.6 роботи.

Контрольні вибірки використовуються в оцінці якості навчання розпізнавального алгоритму, яка розглядається в розділі 3 роботи.

Розкриємо зміст процедури [16, 17], яка використовувалась в роботі для формування з наявного в базі даних [16, 17] запису електрокардіограми пацієнта достатніх за обсягом вибірок сигналів для кожного класу, що розглядається. Вважаємо, що в цій базі даних тип кожного екземпляру сигналу, його належність певному конкретному класу визначений гарантовано правильно [4, 5]. Для спрощення обробки даних призначається однаковим обсяг вибірок для усіх цих класів.

Екземпляри сигналів для контрольних вибірок будуть формуватися аналогічно.

Результатом формування навчальних (і контрольних) вибірок мають бути, таким чином, масиви однакової розмірності екземплярів сигналів кожного класу. При цьому кожний екземпляр сигналу в таких масивах – це послідовність його

відліків в цифровому вигляді. Принцип формування навчальних і контрольних вибірок розкриває рисунок 1.6.

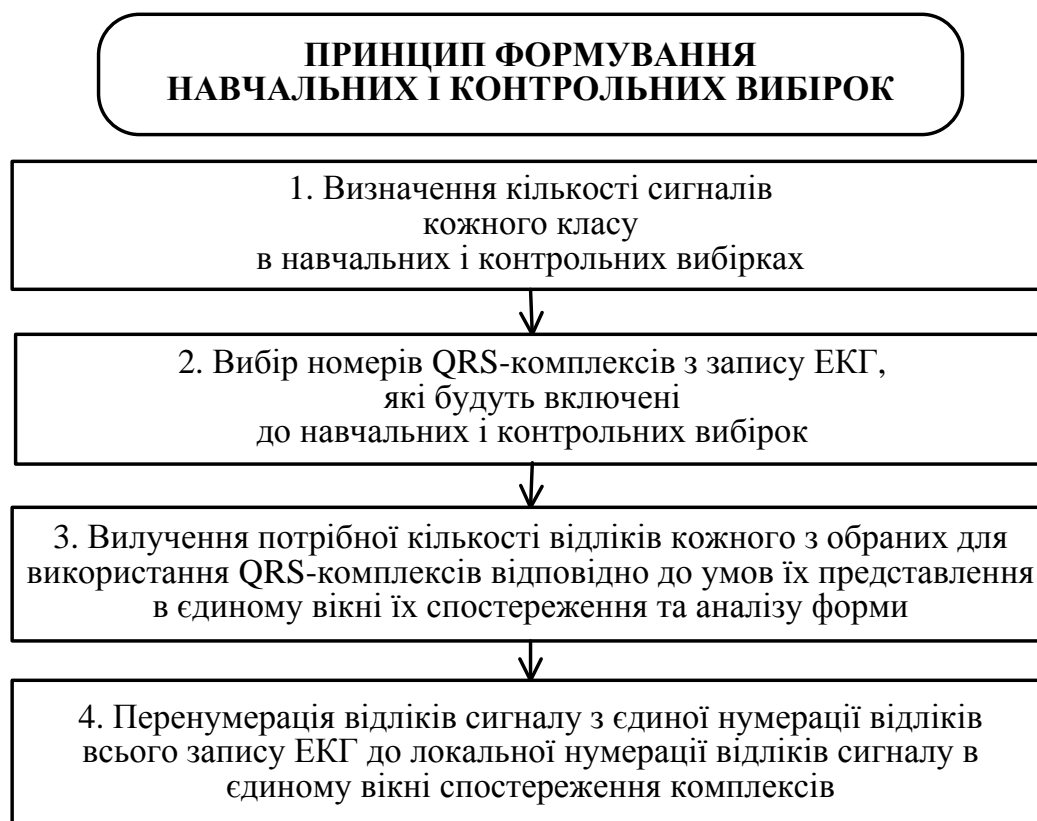


Рисунок 1.6 – Принцип формування навчальних і контрольних вибірок

До цієї графічної характеристики цього принципу додамо, що кількість сигналів у складі навчальної і контрольної вибірки для будь-якого з трьох класів, що розглядаються (QRS-комплекси N, A і V типів), проводиться шляхом перегляду запису відведення ЕКГ та підрахунку наявних екземплярів кожного типу. Знаходиться клас з мінімальною кількістю наявних в запису екземплярів. З міркувань забезпечення однакового обсягу навчальної і контрольної вибірки сигналів кожного класу ця кількість ділиться навпіл і таким призначається обсяг всіх навчальних і контрольних вибірок. У разі непарної мінімальної кількості комплексів в одному з класів в контрольних вибірках буде на один екземпляр сигналів більше.

Стосовно вибору номерів QRS-комплексів для вибірок зазначимо, що в роботі реалізується принцип рівномірності відбору екземплярів по всьому запису для цих вибірок. Для типу комплексів з їх мінімальною наявною кількістю до навчальної і контрольної вибірок потрапляють усі екземпляри. Для класів з "зайвими" комплексами робиться рівномірна вибірка необхідної кількості екземплярів з рівномірним пропуском в запису зайвої кількості комплексів по всіх даних і виключенням їх з розгляду.

Потрібна кількість відліків кожного з QRS-комплексів для їх аналізу в єдиному вікні їх спостереження визначається з таких міркувань. Застосування такого вікна для спостереження та аналізу форми QRS-комплексів, що розпізнаються, забезпечує синхронізацію розгляду цих комплексів під час їх розпізнавання.

Зміст цієї синхронізації, по-перше, полягає в тому, що для будь-якого комплексу в запису розглядається однакова кількість відліків, яка складає ширину єдиного вікна спостереження, тобто ширину фрагменту такого сигналу, по якому визначається його тип, його належність одному з класів.

По-друге, R-пік будь-якого обраного для розгляду QRS-комплексу, відповідно до позначки для нього ЕКГ в базі даних, переноситься з відповідною кількістю відліків сигналу зліва і справа на шкалі часу на одне і те ж саме положення у вікно спостереження сигналів (на відлік № 48 в цьому вікні).

Ліворуч від положення R-піку у вікні спостереження рівномірно розташовано 47 відліків сигналу, праворуч – ще 80, разом у цьому вікні 128 відліків. Всього вікно спостереження охоплює інтервал часу тривалістю приблизно 500 мс.

Ширина єдиного вікна спостереження для розпізнавання типів QRS-комплексів (кількість відліків для аналізу їх форми під час розпізнавання), положення R-піку кожного комплексу у вікні були підібрані емпірично за критерієм найбільшої валідності розпізнавання таких сигналів під час попередніх досліджень розпізнавання типів комплексів в запису ЕКГ пацієнта.

Під час таких досліджень бралися до уваги такі обставини. У вікно спостереження мали потрапляти відліки не більш, ніж одного комплексу, з

урахуванням можливою малості періоду їх повторення у разі високої частоти пульсу у пацієнта. Деяку варіабельність мають форма елементів цих комплексів і їх тривалість. В рамках таких обмежень охоплення в часі вікном спостереження комплексів, що розглядаються, це охоплення має бути максимально можливим для найбільш повного врахування особливостей форми комплексів різних типів під час їх розпізнавання.

Вважалося також, що зайві, малоінформативні відліки сигналів, що потрапляють в єдине вікно, вікно можуть бути в подальшому вилучені з розгляду шляхом застосування спеціальних процедур відбору інформативних ознак сигналів [11, 13, 14].

Слід також зауважити, що в роботі має місце перенумерація відліків сигналу. З єдиної нумерації відліків всього запису ЕКГ вона перетворюється в локальну нумерацію відліків відокремлених фрагментів сигналу в єдиному вікні спостереження комплексів і саме така нумерація використовується для відліків в масивах екземплярів комплексів навчальних і контрольних вибірок. Така перенумерація проводиться одночасно з вибіркою комплексів з запису ЕКГ.

Формування навчальних і контрольних вибірок проводиться по черзі по класах. Розподілі всієї множини відібраних екземплярів комплексів кожного типу між навчальною та контрольною вибірками здійснюється також за принципом чергування: перший комплекс іде до навчальної вибірки, другий – до контрольної і так далі.

Таким чином, завдяки застосованій фрагментації запису та розміщення отриманих фрагментів комплексів в єдиному вікні спостереження в умовах нерегулярності, варіабельності періоду їх повторення в запису, а також деякої варіабельності форми елементів і їх тривалості усі комплекси ставляться в однакові умови для прийняття рішень про їх типи під час розпізнавання.

Отримані таким чином вибірки сигналів (QRS-комплексів) вважаються первинними. Надалі відліки цих сигналів в такому початковому вигляді перераховуються у відліки характеристик їх форми, які використовуються як

ознаки для розпізнавання сигналів та формування еталонів для їх класів, про що йдеться далі.

1.4 Ознаки сигналів для їх розпізнавання і зміст попередньої обробки даних для їх розрахунку.

Як вже зазначалося, за ознаки для навчання розпізнавальних алгоритмів та подальшого розпізнавання сигналів контрольної вибірки взято відліки сигналів, що перераховані у відліки характеристики їх форми.

Саме перехід до характеристики форми сигналів і є змістом їх попередньої обробки, що є найпершою операцією у алгоритмі розпізнавання поточних сигналів та підготовки вибірок.

Попередня обробка сигналів – це процедура, що застосовується при формуванні контрольних та навчальних вибірок, еталонних сигналів, тощо, для підвищення вірності розпізнавання та прийняття рішень про тип вхідної реалізації. Зміст попередньої обробки зводиться до приведення сигналів до характеристики форми. [3, 8, 5, 6]

Характеристикою форми вважається зображення сигналу, з якого вилучено постійну складову, а сумарна енергія усіх його відліків приведена до одиничного значення.

З математичної точки зору процедура приведення сигналу до характеристики форми має такий вигляд.

$$u = x - \frac{1}{N} \sum_{i=1}^N x_i, \quad (1.1)$$

де x – вхідний сигнал з первинного набору даних;

N – кількість відліків вхідного сигналу.

$$u_i^H = \frac{u_i}{\sqrt{\sum_{i=1}^N u_i^2}}, \quad (1.2)$$

де u_i – i -тий відлік сигналу з вилученою постійною складовою, $i = 1:N$.

За формулою (1.1) оцінюється постійна складова сигналу, яка вилучається з кожного з відліків x_i їх первинної послідовності. У формулі (1.2) відбувається нормування відліків сигналу, тобто їх масштабування – приведення сумарної енергії кожного з сигналів до одиничного значення. [3, 6]

Отже, таким чином, було розглянуто процедуру попередньої обробки сигналів, змістом якої є приведення сигналів навчальних та контрольних вибірок до характеристики їх форми. Попередню обробку проходять як кожна зі вхідних реалізацій сигналу, що поступає до розпізнавального алгоритму, так і сигнали навчальної вибірки, що використовуються у навчанні алгоритму. Процедури обробки сигналів аналогічні як для базових процедур, так і для розробленої процедури комбінованого навчання, зміст якого розкривається у наступному підрозділі.

1.5 Принцип і порядок комбінованого навчання розпізнавальних алгоритмів

Основні переваги та недоліки зазначених процедур навчання з учителем та самонавчання розкриваються наступними положеннями.

1. Основною перевагою процедури навчання з учителем є збереження традиційної медичної класифікації сигналів. Недоліком такого методу є використання лише одного еталону для кожного з заданих учителем класів.

2. Перевага алгоритму самонавчання – можливість більш детального розбиття сигналів на групи (кластери) за спільними ознаками, але основним недоліком є відсутність класифікації в алгоритмі, тобто сигнали групуються лише

за подібністю, незалежно від класу. Такий недолік ускладнює повноцінне застосування алгоритмів самонавчання.

Під комбінуванням (рис. 1.7) розуміється поєднання процедури навчання з учителем із внутрішньою кластеризацією, що притаманна процедурам



Рисунок 1.7 – Загальний зміст процедури комбінування в машинному навчанні розпізнавальних алгоритмів

самонавчання, в межах кожного типу заданого вчителем [27-29, 5].

Основна мета введення такого типу навчання – підвищення кількості локальних еталонів в межах кожного класу, що повинно стати передумовою більш детального врахування розподілу зображуючих сигнали точок і відповідно підвищити вірність розпізнавання вхідних сигналів із збереженням традиційної класифікації [8]. Загальна схема такої процедури комбінованого навчання зображена на рисунку 1.8.

Яскраво проілюструвати зміст роботи запропонованого алгоритму можна за допомогою деякого скупчення точок (рис. 1.9 (а)), причому деякі з точок віддалені від основного скупчення та є частковими випадками. При навчанні з учителем ним

задається класифікація на групи, в ілюстрованому випадку точки поділяються на 3 класи (рис. 1.9 (б)), відповідно для кожного класу утворюється один еталон (рис. 1.9 (в)), що є усередненим значенням з усіх елементів певного класу навчальної вибірки.

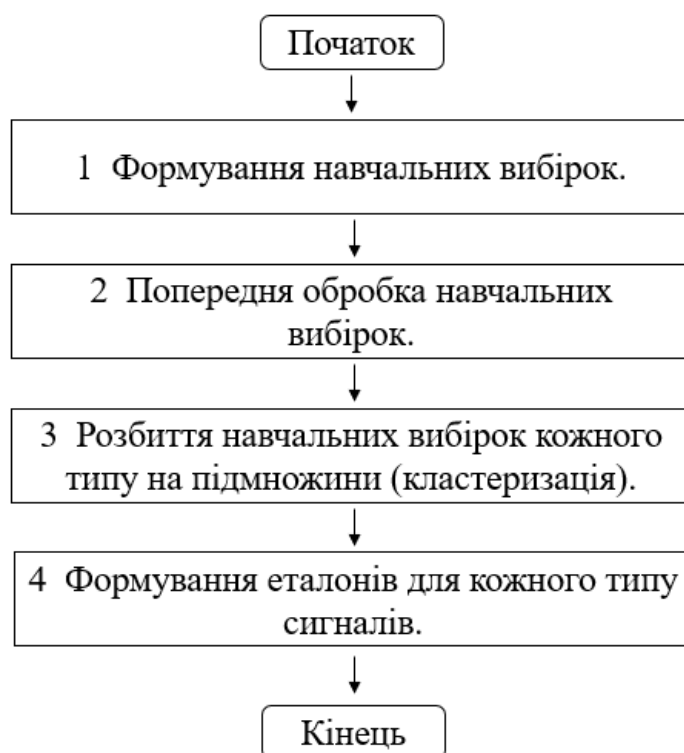


Рисунок 1.8 – Блок-схема процедури комбінованого навчання алгоритму розпізнавання

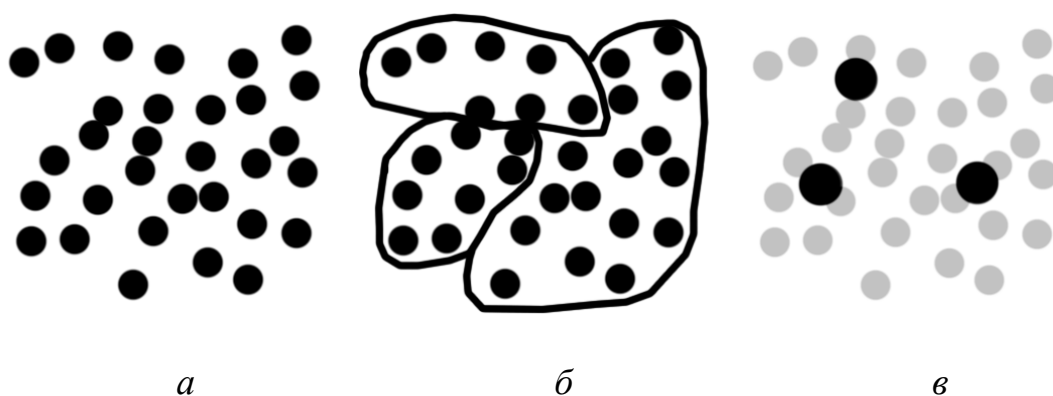


Рисунок 1.9 – Ілюстрація роботи алгоритму навчання з учителем на прикладі деякої групи точок, де *a* – навчальна вибірка; *б* – навчальна вибірка з класифікацією заданою вчителем; *в* – ймовірні еталонні зразки сигналів, утворені за допомогою усереднення всіх реалізацій навчальної вибірки в межах класу.

На рисунку показано, що приблизна локація еталонів – серед основного скупчення точок.

Здійснивши внутрішню кластеризацію навчальних вибірок, в межах кожного з заданих типів, отримаємо підгрупи подібних між собою реалізацій. [3, 7-10] Еталони з новостворених груп утворюються способом, аналогічним базовому алгоритмі – усередненням. Таким чином отримаємо деяку кількість еталонів для кожного з класів (рис. 1.10).

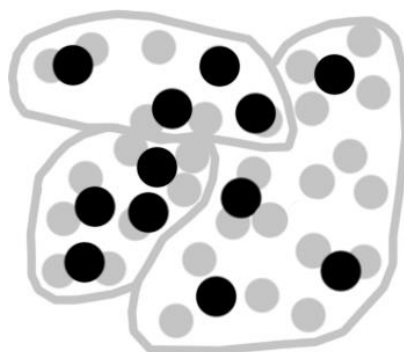


Рисунок 1.10 – Ілюстрація ймовірного результату роботи алгоритму комбінованого навчання – кожна група має декілька локальних еталонів (кластерів)

Таким чином утворилася більш розгалужена система еталонів, яка охоплює більшість часткових випадків навчальних вибірок в межах окремих класів. В такому випадку розпізнавання поточної реалізації повинне проводитися в два етапи. Перший – визначення найбільш схожого еталона (кластера) у всій групі навчальних вибірок, другий – визначення належності найбільш схожого еталона до заданих учителем класів і присвоєння типу за нею.

Наведена ілюстрація підтверджує сприяння методу комбінованого машинного навчання алгоритмів розпізнавання збільшенню кількості вірно класифікованих рішень.

Процедура комбінування навчання потребує, таким чином, вирішення питання кластеризації навчальних вибірок.

Кластерний аналіз – це комплекс статистичних процедур, що вирішує задачу розбиття заданої вибірки даних та упорядкування підмножин (кластерів). Перелік кластерів визначається в процесі виконання процедури кластеризації і початково не задається. Завдання кластеризації відноситься до статистичної обробки даних. [30]

В цілому, кластерний аналіз може бути застосований для вирішення таких задач.

1. Групування об'єктів вибірки за їх подібністю або іншою мірою схожості.
2. Стиснення даних – скорочення кількості елементів вхідної вибірки.
3. Розробка типології та класифікації.
4. Перевірка наявності відомих груп даних в наданих вибірках.

Загалом, кластерний аналіз може бути застосовуватися для безлічі задач, а тому наведений перелік можна продовжувати. Кластерний аналіз широко розповсюджений в багатьох сферах наукової діяльності, таких як економіка, соціологія, біологія, біомедицина, інформатика тощо. Особлива увага до нього привертається у в сфері машинного навчання – задача кластеризації даних належить до алгоритмів самонавчання і є інструментом для автоматичного поділу та упорядкування наявної вибірки даних за їх подібністю, або іншою мірою подібності, що задається дослідником [27, 30-32].

Загальна процедура кластеризації вибірки даних, не залежно від типу досліджуваного об'єкту складається з наступних кроків.

Крок 1. Відбір та підготовка вибірки до кластеризації – попередня обробка вхідних даних та виділення векторів їх характеристик, за якими проводитиметься кластеризація.

Крок 2. Вибір метрики та характеристик за якими буде проводитись оцінка об'єктів вибірки. Об'єкти вибірки можуть описуватися як відстанями до усіх інших об'єктів простору, так і за ступенем схожості об'єкта з іншими об'єктами у просторі вибірки. Здебільше до метрик відносять відстань між елементами кластеру, їх

центроїдами — центрами мас кластерів, тощо. Метрика обирається в залежності від особливостей характеристик досліджуваного простору, розташування точок та від вимог до утворюваних кластерів – кластери можуть суттєво відрізнятися в залежності від обраної метрики, задача вибору якої повністю лежить на досліднику.

В алгоритмах кластеризації найчастіше зустрічається застосування Евклідової відстані та її квадрата. Відстань Евкліда (1.3) та її квадрат (1.4) є найбільш простою, застосованою та найбільш зрозумілою метрикою.

$$p(x, x') = \sqrt{\sum_i^n (x_i - x'_i)^2}, \quad (1.3)$$

$$p(x, x') = \sum_i^n (x_i - x'_i)^2, \quad (1.4)$$

де $p(x, x')$ – відстань між об'єктами x, x' ;

i – номер складової розкладу

n – кількість складових розкладу цих об'єктів;

x_i, x'_i – координати об'єктів вибірки.

Використовувані в роботі метрики для кластеризації навчальних вибірок визначена в завданні на її виконання. Квадрат Евклідової відстані застосовується для більш детального урахування віддалених об'єктів.

Як метрика кластеризації при тестовій перевірці алгоритму комбінованого навчання буде застосовуватися Евклідова метрика, так як вона є найпростішою в реалізації та не потребує додаткових затрат машинного часу та обчислювальної потужності. [27, 30-32]

Для порівняння впливу вибору метрики кластеризації на кінцевий результат розпізнавання, окрім Евклідової метрики, буде проведена кластеризація за кореляційною відстанню. [8, 9]

$$\alpha_z = 1 - z^*, \quad (1.5)$$

$$\text{де } z^* = \frac{1}{2} (1 + z);$$

$$z = \sum_{i=1}^M p_i q_i;$$

p_i, q_i – відліки характеристик форми сигналів вибірки, за якими розраховується їх скалярний добуток z як кількісний показник близькості їх форми;

M – кількість відліків кожного сигналу вибірки.

За своїм змістом показник α_z є показником відмінності форми сигналів, своєрідною відстанню між характеристиками їх форми. Початкова шкала $[-1;1]$ значень скалярного добутку, який розглядається показник близькості форми сигналів, спочатку перераховується до більш зручної тут шкали $[0;1]$. Тобто різниця $\alpha_z = 1 - z^*$ доповнює наявну схожість z^* до максимально можливого одиничного значення і припадає на нестачу співпадіння форми сигналів, іншими словами, на відмінність їх форми.

Інші найбільш відомі метрики [30, 31], такі, як відстань Хеммінга (середня різниця по координатам точок вибірки), Чебишева (максимум модуля різниць координат вибірки), косинусна відстань (одиниця мінус косинус кута векторів між точками) в роботі не розглядаються.

Критерій α_z , що розглядається в роботі, є найбільш близьким за своїм змістом до випадку використання косинусної відстані, з урахуванням наведеного визначення характеристики форми сигналу.

Крок 3. Обчислення матриць відстані, або подібності враховуючи усі елементи вибірки.

Крок 4. За обраним методом кластерного аналізу, на основі матриць подібності відбувається утворення кластерів. Метод кластеризації обирається дослідником за рядом факторів: особливості простору і розподілу точок, об'єму вибірки, необхідного класу точності, оптимального використання ресурсів, швидкодії, тощо.

Крок 5. Візуалізація результатів кластеризації.

Як вже зазначалося, задача вибору методу кластеризації лежить на досліднику. Однозначної класифікації методів кластеризації не існує, а їх кількість

є незліченною. Загалом алгоритми кластеризації можна класифікувати за їх принципом роботи та підходом до визначення кластерів із даних вибірки:

- Ієрархічні – характеризуються послідовним об'єднанням менших елементів у більші і скороченні їх кількості при цьому. На виході такого алгоритму виникає ієрархія кластерів – дендрограма. ;
- Ітеративні – робота базується на виконанні ітеративного процесу, що вирішує завдання оптимізації цільової функції, визначаючої якість кластеризації множини об'єктів;
- Групові;
- Алгоритми з використанням нейронних мереж та штучного інтелекту;
- Алгоритми засновані на теорії графів;
- Алгоритми засновані на логічному підході;
- Алгоритми засновані на імовірнісному підході.

Через високу кількість алгоритмів кластеризації описаний перелік можна продовжувати.

Формально задача кластеризації описується таким чином. Дана вибірка даних $I = \{i_1, i_2, \dots, i_j, \dots, i_n\}$, де $i_j = \{x_1, x_2, \dots, x_h, \dots, x_m\}$ – елемент вибірки що характеризується набором параметрів. Змістом кластеризації в даному випадку є побудова підмножини: $C = \{c_1, c_2, \dots, c_k, \dots, c_g\}$, де $c_k = \{i_b, i_p / i_l \in I, i_p \in I, p(i_b, i_p) < \sigma\}$ – кластер, який має в собі подібні об'єкти з множини I . σ – величина, що визначає міру близькості для включення об'єктів в один кластер, $p(i_b, i_p)$ – відповідно міра близькості. [6, 8, 9, 27-29]

Рішення задачі кластеризації по ряду причин є принципово неоднозначним.

1. Кожен з критеріїв оцінки якості кластеризації даватиме різний результат, а в деяких випадках такі параметри взагалі не передбачаються методами. Однозначно найкращого критерію оцінки якості не існує.

2. Результат кластеризації напряму залежить від обраної метрики та числа кластерів, тип та кількість яких задаються дослідником, ґрунтуючись лише на суб'єктивній оцінці умов кластеризації.

В роботі застосовується алгоритм кластеризації методом k -середніх. Цей метод виділяється популярністю, швидкістю роботи та простотою реалізації. Такі переваги є досить суттєвими для використання у комбінованому навчанні алгоритму розпізнавання, так як додаткова кластеризація ускладнює стандартний алгоритм і є сенс скористатися можливістю мінімізувати це ускладнення.

Метод k -середніх – це алгоритм кластеризації, змістом якого є поділ вибірки на k кластерів, щоб кожна реалізація вибірки належала до кластера з найближчим до нього середнім значенням. Метод відноситься до класу алгоритмів квадратичної помилки і є ітеративним.

Таким чином, після кластеризації вибірки сигналів методом k -середніх в кластери збираються сигнали, що найближчі того ж самого центроїда, тобто концентруються в групи близькі за формою сигнали. Тобто кластери охоплюють скупчення близьких між собою сигналів. Такий алгоритм як раз і враховує особливості локації сигналів навчальних вибірок в класах використовуваної класифікації, що й потрібно в роботі. Центроїди отриманих кластерів можливо використовувати в якості систем локальних еталонів в класах сигналів, що розпізнаються.

Алгоритми такого типу розглядають задачу кластеризації, як задачу оптимального розбиття об'єктів на кластери. Критерій оптимальності такого алгоритму визначається за формулою (1.6). [27]

$$V = \sum_{i=1}^K \sum_{x \in S_i} (x - \mu_i)^2, \quad (1.6)$$

де S_i – кластер;

μ_i – центроїд кластера S_i ;

k – кількість кластерів.

Процедура кластеризації алгоритмом k -середніх (рис. 1.11) складається з наступних етапів.

1. Дослідником задається необхідна кількість кластерів k . У просторі вибірки випадково утворюється k точок, що є початковими центроїдами кластерів.

2. Кожна реалізація вибірки відноситься до центроїда за мірою схожості (найчастіше за відстанню Евкліда – найближчі центроїди), утворюючи кластери.

3. Перерахунок центроїдів новоутворених кластерів, матриця відстаней оновлюється. Обчислення відбувається за формулою:

$$\mu_l^{(i)} = \frac{\sum_{j=1}^d u_{ij} m_j}{\sum_{j=1}^d u_{ij}}, 1 \leq i \leq k, \quad (1.7)$$

де $\mu_l^{(i)}$ – центроїд кластеру;

d – розмірність досліджуваних об'єктів;

k – кількість кластерів;

l – номер ітерації;

u – матриця розбиття;

m – досліджуваний об'єкт.

4. Повторення процедури починаючи з кроку 2, до досягнення кількості ітерацій, при якій кластерні центри стануть стійкими.

Незважаючи на простоту та високу швидкість обчислень – алгоритм містить свої недоліки, основним з них є чутливість до викидів точок, що присутні у медико-біологічних сигналах.

Результат роботи алгоритму такого типу [27, 29, 31] проілюстровано на рисунку 1.11.

Для методу k -середніх наявна велика кількість додаткових модифікацій та розширень, що можуть оптимізувати роботу алгоритму. Одна з них, наприклад, здійснює пошук початкових центроїдів, замість утворення їх випадковим чином. Найвідомішою з таких модифікацій є так званий алгоритм C -середніх. На відміну від первісного алгоритму в ньому розраховуються не відстані до центроїду кластеру, а ступінь належності до нього за мірою схожості. Модифікації подібного роду створені для зменшення неоднозначності результатів кластеризації і застосовуються при великих об'ємах вибірки даних. [31]

Метод k -середніх досить широко використовується в алгоритмах самонавчання, причому не завжди за прямим його призначенням – створення класифікації розбивкою на кластери.

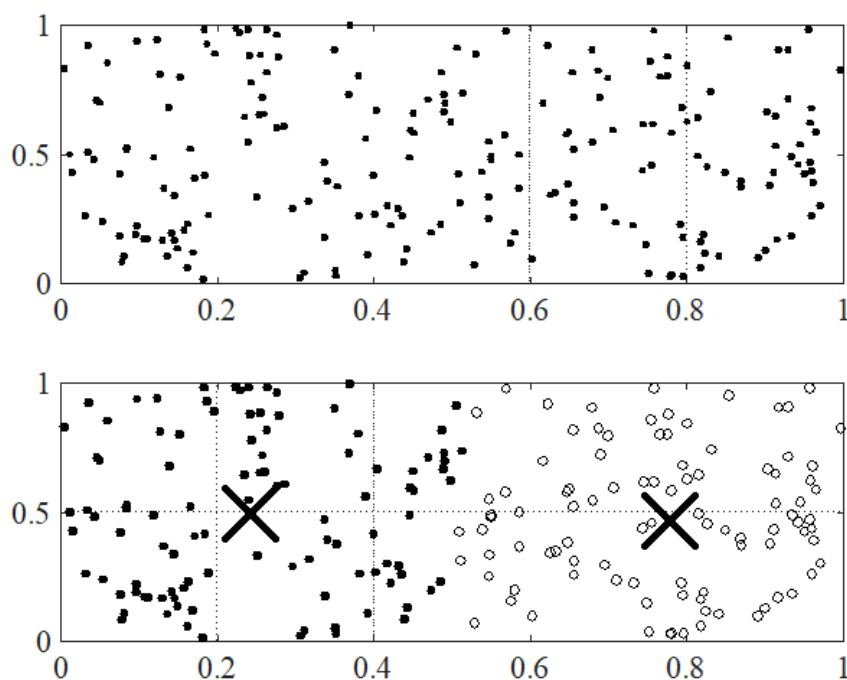


Рисунок 1.11 – Ілюстрація результатів кластеризації
методом k-середніх на 2 кластери

Як вже зазначалося, існує досить велика кількість алгоритмів кластеризації, не маючих однозначної класифікації. Нижче наведено короткий перелік таких алгоритмів відповідно до перерахованих вище підходів.

Алгоритми засновані на імовірнісному підході:

- k-середніх;
- k-медіан;
- FOREL;
- ЕМ-алгоритм.

Алгоритми з використанням нейронних мереж та штучного інтелекту:

- Алгоритм нечіткої кластеризації С-середніх;
- Алгоритм з використанням карт Кохенена;
- Генетичні алгоритми.

Кожен з описаних методів має свої переваги. [27, 31] Ієрархічні методи відрізняються від інших найточнішим розбиттям на кластери, нейромережі відрізняються своєю паралелізацією і відповідно швидкодією. Алгоритм k-середніх відрізняється від інших своєю простотою та відносно малою ресурсозатратністю.

На практиці частково застосовуються групові, або гібридні методи, які поєднують у собі операції, взяті із різних методів з урахуванням їх властивостей. [30-32]

Отже, даний розділ розкриває процес створення, мету та зміст комбінованого навчання алгоритму розпізнавання, з чого можна відмітити такі положення.

1. Мета комбінування – підвищення вірності розпізнавання медико-біологічних сигналів за рахунок більш детального урахування особливостей розподілу точок відліків реалізацій в межах відомих класів.

Зміст комбінування – додаткове розбиття навчальних вибірок на кластери, утворення декількох характерних еталонів в межах кожного класу що характеризуватимуть наявні особливості розподілу точок, зі збереженням традиційної класифікації при цьому.

2. Ілюстрації роботи алгоритму комбінованого навчання наводяться для підтвердження прогнозу спроможності запропонованого алгоритму до розкриття резервів у підвищенні вірності розпізнавання сигналів.

3. Попередня обробка сигналів та принцип прийняття рішень про тип вхідної реалізації – аналогічні тим, що властиві навчанню з учителем.

4. Кластерний аналіз – широко розповсюджений інструмент математичної статистики, що зустрічається в багатьох сферах наукової діяльності, а особливо в аналізі медичних даних у діагностичних системах. Особлива увага приділяється кластеризації вибірок сигналів в області машинного навчання розпізнавальних алгоритмів, а саме в побудові алгоритмів самонавчання.

5. Одним із застосувань кластерного аналізу можна скористатися для розробки комбінованого алгоритму – групування даних, тобто розбиття множини даних на підмножини за критерієм їх близькості один до одного.

6. Алгоритм кластеризації методом k-середніх є одним з найпростіших варіантів розбиття вибірки на кластери, кількість яких необхідно задавати наперед. Ітеративність даного методу забезпечує чітку та однозначну кластеризацію

вибірок. Для кластеризації вибірок при комбінованому навчанні цей алгоритм підходить щонайбільше, оскільки його простота та надійність гарантує результат з невеликими затратами ресурсів.

7. Процедура кластеризації методом k-середніх передбачає можливість обрання метрики для оцінки відстані між точками вибірки, що додає простору для досліджень, адже застосовувана метрика впливає на результат кластеризації. Базовою метрикою обрано найпростіший варіант – відстань Евкліда.

8. Алгоритм кластеризації методом k-середніх реалізований в багатьох програмних середовищах, що дає можливість візуалізувати результат на кожному етапі роботи алгоритму.

9. На виході алгоритму маємо центроїди кластерів, що являють собою середнє значення усіх елементів кожного з них, тобто і є локальними еталонами, які пропонується застосовувати в алгоритмі комбінованого навчання.

У наступному підрозділі розкриваються загальні висновки, на основі описаних положень.

1.6 Висновки до розділу 1

Основними є такі результати даного розділу.

1. Розкрито зміст основних класичних алгоритмів машинного навчання, визначено їх переваги та недоліки, що є передумовами до створення комбінованого алгоритму.

2. Проаналізовано мету та зміст комбінування алгоритму навчання з учителем із додатковою кластеризацією вибірок, що притаманна алгоритмам самонавчання. Графічно проілюстровано принцип комбінування у алгоритмі навчання. З чого визначено такі положення:

- для створення комбінованого навчання відібрано алгоритм навчання з учителем, що працюватиме в комбінації з додатковою кластеризацією навчальних вибірок в межах заданих учителем типів;
- комбіноване навчання алгоритму розпізнавання призведе до появи локальних еталонів в межах кожного типу, на відміну від алгоритму навчання з учителем, де як в ролі еталона використовувався середній вигляд усіх реалізацій заданого типу;
- метою реалізації алгоритму комбінованого навчання є створення потужного та універсального інструментарію для розпізнавання та подальшої класифікації медико-біологічних сигналів з високою достовірністю правильних рішень.

3. Запропоновано порядок прийняття рішень у разі комбінованого навчання. Застосування широковідомих алгоритмів додає універсальності застосування розробленого методу.

4. Розкрито основний зміст та логіку роботи процедури кластеризації. Проведено короткий огляд алгоритмів кластеризації, метрик та їх характерних особливостей. Для подальшої розробки програмного інструментарію обрано алгоритм кластеризації методом k-середніх із базовою Евклідовою метрикою, як найпростіший, найменш затратний та легкий до реалізації метод. Для порівняння та оцінки впливу вибору метрики на результати кластеризації паралельно з Евклідовою метрикою буде застосовуватися розрахунок кореляційних відстаней.

Здобуті результати будуть використані в подальшій розробці програмних процедур комбінованого машинного навчання алгоритму розпізнавання.

З аналізу змісту завдання комбінованого навчання алгоритмів розпізнавання медико-біологічних сигналів, а також відповідно до наведених рішень про зміст основних обчислювальних процедур обробки даних щодо виконання цих завдань і самого розпізнавального алгоритму, який використовується для перевірки результативності його комбінованого навчання, впливають такі основні висновки.

1. В основу комбінованого машинного навчання процедур розпізнавання медико-біологічних сигналів з метою підвищення його якості можливо покласти

більш детальне врахування розподілів сигналів по вірогідності в їх класах з медичною інтерпретацією шляхом додаткового розбиття навчальних вибірок за допомогою інструментарію кластеризації на підмножини в локаціях скупчення зображуючих сигнали точок та використання багато еталонних описів класів сигналів під час їх розпізнавання.

2. Завдання комбінованого навчання процедур розпізнавання медико-біологічних сигналів є комплексним і передбачає не тільки розробку процедур і алгоритму такого навчання, але й підтвердження результативності їх роботи, для чого має бути конкретизований вид і формат медико-біологічних сигналів, що розпізнаються, підібраний для проведення розробок сам розпізнавальний алгоритм, що навчається з учителем, проведене його комбіноване навчання, здійснені статистичні розрахунки чутливості, специфічності, загальної валідності розпізнавання сигналів цим алгоритмом після такого навчання на контрольних вибірках у порівнянні з обробкою даних з використанням одне еталонних описів навчальних вибірок за звичайною схемою навчання з учителем.

3. Загальна схема реалізації і перевірки результативності комбінованого навчання розпізнавального алгоритму має складатися з конкретизації типів медико-біологічних сигналів, що розпізнаються, розкриття їх класифікації і системи ознак для розпізнавання, а також з процедур формування навчальних і контрольних вибірок, підготовки опису сигналів в їх складі на мові ознак для розпізнавання, вибору розпізнавального алгоритму, його навчання комбінованим і звичайним способами і проведення потрібної порівняльної статистичної оцінки показників якості розпізнавання сигналів на тестовому прикладі.

4. Сигнали для проведення досліджень комбінованого навчання розпізнавальних алгоритмів доцільно підбирати в цій роботі такими, щоб вони мали конкретну медично інтерпретовану класифікацію, гарантовано визначену приналежність кожного з їх екземплярів певному класу, бажано щоб вони були процесами в часі та мали достатню значну для їх розпізнавання стабільність форми для їх розпізнавання за нею і деяку її варіабельність, що зазвичай має місце для медико-біологічних сигналів. В цілях цієї роботи припустимо використання

сигналів в їх запису за результатами цифрових вимірювань в форматі послідовності відліків на рівномірній сітці часу на достатньо високій частоті знімання даних, щоб забезпечувати якісне відображення їх форми. Важлива також достатня кількість цих сигналів для формування навчальних і контрольних вибірок.

5. Простими і зручними для опису форми сигналів є послідовності відліків характеристики їх форми, які є тими ж самими відліками сигналів в початковому вигляді, з яких видалена їх постійна складова і які потім спеціальним чином пронормовані по інтенсивності, завдяки чому поліпшуються умови розпізнавання сигналів, зокрема зменшується вплив нерегулярних коливань ізолінії сигналу ЕКГ пацієнта.

6. Для проведення розробок і досліджень щодо комбінованого навчання процедур розпізнавання медико-біологічних сигналів, для розкриття змісту такої виконуваної обробки даних цілком підходить використання півгодинного запису електрокардіограми пацієнта з бази даних в Інтернет з розміткою типів усіх QRS-комплексів і положення R-піків кожного з них в запису з розумінням під класами сигналів, що розпізнаються, N, A і V типи таких комплексів за умови їх спостереження і аналізу в єдиному часовому вікні, з огляду на змінність частоти їх повторення та деякої варіабельності форми. Спрощення оцінки результативності комбінованого навчання розпізнавальних алгоритмів забезпечується у разі використання навчальних і контрольних вибірок таких сигналів у разі однаковості їх обсягу.

7. Про результативність комбінованого навчання розпізнавальних алгоритмів в порівнянні з традиційним навчанням за схемою навчання з учителем достатньо судити за показниками чутливості, специфічності і загальної валідності розпізнавання сигналів досліджуваними алгоритмами після їх навчання розробленими процедурами.

8. В цілому, комбіноване машинне навчання процедур розпізнавання медико-біологічних сигналів доцільно розглядати як завдання удосконалення програмного забезпечення, яке призначене для розкриття резервів підвищення показників

вірності розпізнавання медико-біологічних сигналів в системах діагностики пацієнтів.

Алгоритмічна і програмна реалізація наведених в розділі 1 процедур у складі комплексного завдання комбінованого навчання алгоритмів розпізнавання медико-біологічних сигналів представлені в наступному розділі роботи.

2 ПРОГРАМНІ ПРОЦЕДУРИ АЛГОРИТМУ РОЗПІЗНАВАННЯ СИГНАЛІВ ІЗ ЙОГО КОМБІНОВАНИМ НАВЧАННЯМ

2.1 Процедура формування навчальних і контрольних вибірок сигналів

Вибірки відбираються з генеральної сукупності даних за допомогою заданих дослідником процедур, відповідно до чого постає питання про алгоритми формування вибірок. [7, 32]

Вибірки даних, що застосовуються у процедурах машинного навчання, поділяються на два типи – контрольні та навчальні.

Під процесом формування вибірки (рис. 2.1) в роботі розуміється процес реалізації процедури, що включає в себе отримання вхідних даних, наданих вчителем, формування навчальних та контрольних вибірок по типах сигналів із забезпеченням їх необхідного формату та обсягів.



Рисунок 2.1 – Блок-схема повного процесу формування вибірок для навчання розпізнавального алгоритму та оцінки його якості

Для створення вибірок з безперервного запису електрокардіограми пацієнта з нього вирізаються фрагменти даних на певному проміжку значень в часі, який будемо називати вікном. У випадку QRS-комплексів пацієнта, знятих на частоті 257 Гц, вікно складається із 128 відліків сигналу. Значення ширини вікна обирається таким, при якому вікном охоплюватимуться усі інформативні точки та ознаки комплексу, але при цьому не відбудуватиметься накладання сусідніх вікон.

Вибір положення вікна на запису ЕКГ та синхронізація сигналу у ньому відбувається за відомими координатами R-піків QRS-комплексів і здійснюється таким чином, щоб кожний цей пік припадав у вікні на 48-й відлік.

Більш-детально процес формування вибірок було розкрито у підрозділі 1.3.

Розрізняється досить велика кількість способів формування вибірок. У випадку комбінованого алгоритму навчання, синхронізовані у вікні 128 відліків сигнали відбираються до вибірок таким способом – об'єкти обираються з запису рівномірно через потрібні інтервали із розрахунку на те, щоб їх була задана кількість. Для класу сигналів з мінімальною кількістю QRS-комплексів їх кількість не скорочується.

Отже, в даному підрозділі розглянуто зміст програмної процедура утворення вибірок та підготовки даних для їх застосування у процедурі навчання розпізнавального алгоритму та для оцінки результативності його роботи.

Таким чином, в цілому.

. Розкрито зміст програмної реалізації порядку вибірки сигналів з їх генеральної сукупності. Інтервал відбору обрано відповідно до обсягу вибірки в цілому та обсягів сукупностей QRS-комплексів для кожного із заданих учителем типів.

2. Вибірки сигналів для комбінованого навчання алгоритмів розпізнавання мають однакові обсяги, що спрощує відповідні розрахунки. Обсяг вибірок обрано за найменшою кількістю зразків одного з типів таких сигналів. Кожний відібраний сигнал в своєму початковому вигляді являє собою послідовність його відліків в мілівольтах на рівномірній сітці часу.

Лістинг робочої програми, що виконує завдання формування вибірок із неперервного запису ЕКГ пацієнта наведено у додатку А.

В наступному пункті роботи розкриваються зміст програмних процедур попередньої обробки відібраних фрагментів запису сигналів, які забезпечують отримання їх описів у вигляді послідовностей ознак для розпізнавання.

2.2 Процедури попередньої обробки вхідних сигналів

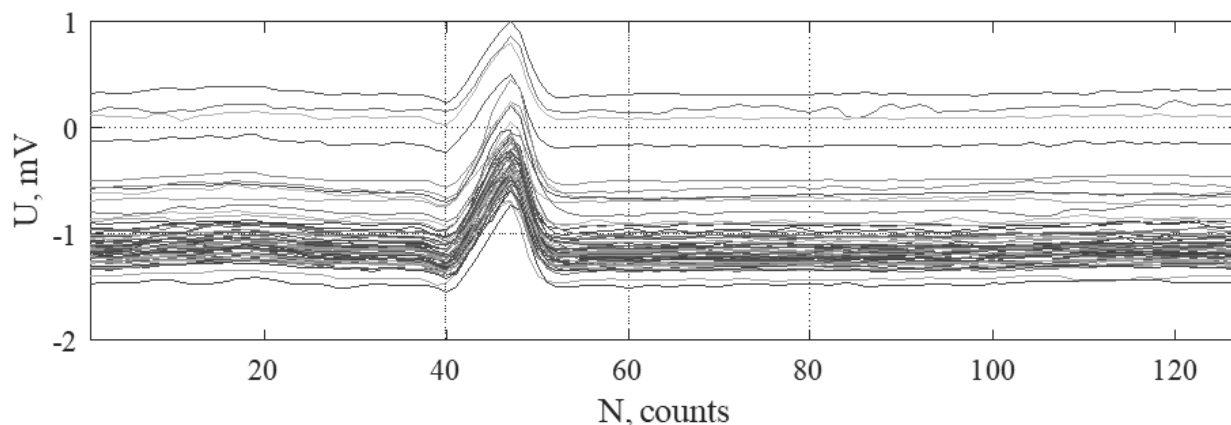
Під попередньою обробкою сигналів розуміється приведення зразків сигналів до характеристики форми [3], тобто вилучення з них постійної складової та приведення сумарної енергії кожного з відліків зразка до одиничного значення (рис. 2.2).



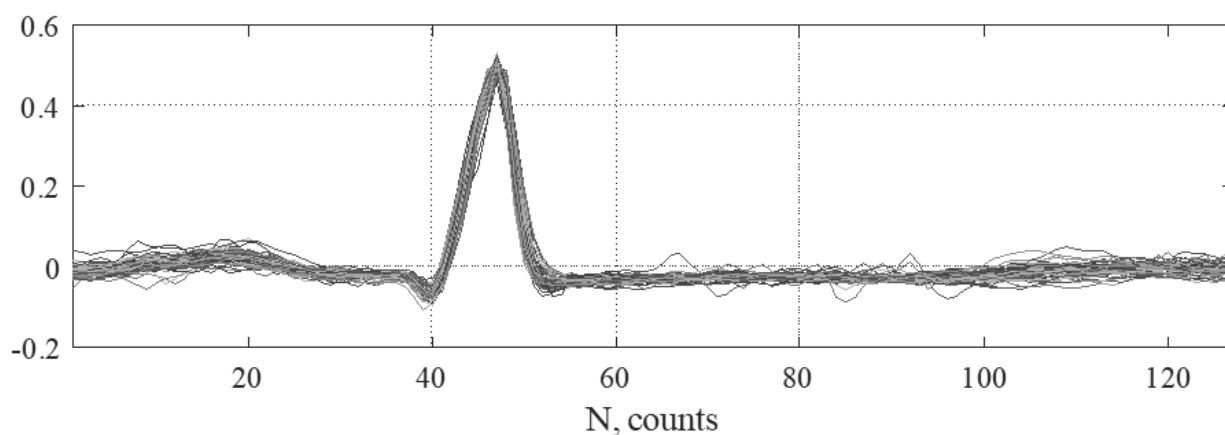
Рисунок 2.2 – Блок схема процедури попередньої обробки сигналів

З теоретичної та математичної сторони процес попередньої обробки сигналів детально розглядався і був обґрунтований у підрозділі 1.4 роботи.

Після проходження попередньої обробки отримуються вибірки характеристик форми сигналів (рис. 2.3).



a



б

Рисунок 2.3 – Ілюстрація результату виконання попередньої обробки на прикладі вибірки QRS-комплексів N-типу з сигналу ЕКГ: *a* – до обробки; *б* – характеристика форми сигналів вибірки.

На рисунку чітко простежується доцільність застосування такої попередньої обробки сигналів.

Так як умовою формування вибірок є рівномірне обрання реалізацій по всьому запису ЕКГ, виникають нерівномірності постійної складової (положення ізолінії), тому варіантом «стабілізації» форми сигналу (співвідношення значень його відліків) вибрано повне її виключення.

Отже, результатами розробки даного підрозділу є визначений зміст та розкриття доцільності застосування процедури попередньої обробки сигналів такого змісту та загальний програмно реалізований алгоритм її проведення.

Ілюстрація результатів виконання процедури попередньої обробки досить чітко показує її вплив на формування еталонів та подальше розпізнавання комплексів з їх використанням.

2.3 Процедури кластеризації навчальних вибірок та формування еталонів

Зазвичай, завдання навчальних алгоритмів полягає в тому, щоб з навчальних вибірок сформувати систему еталонів, які після навчання будуть використовуватися розпізнавальними алгоритмами як апіорна інформація для опису класів сигналів для визначення типів сигналів, які надходять на вхід.

Для процедур розпізнавання сигналів еталонами є характерні зразки сигналів необхідних розпізнаваних типів. У випадку навчання з учителем, кожен з заданих типів представлений одним еталоном, котрий представляє собою середній вигляд характеристики форми всіх реалізацій даного класу з навчальної вибірки. [3, 6]

Як вже зазначалося, процедура комбінування базується на введення додаткової кластеризації в межах заданих учителем типів. Детально процедура кластеризації та обраний для алгоритму метод розглядалася у підрозділі 1.5.

Загальна блок-схема процедури кластеризації наведена на рисунку 2.4.

Найпростішим та найпопулярнішим методом кластеризації, що підходить під вимоги комбінованого алгоритму є метод k -середніх (рис. 2.5).

Змістом кластеризації навчальних вибірок є утворення угруповань з найбільш схожих реалізацій навчальної вибірки для певного класу – кластерів.

Механізм утворення еталонів у такому випадку не відрізнятиметься від аналогічної процедури для навчання з учителем, зокрема еталони являтимуть

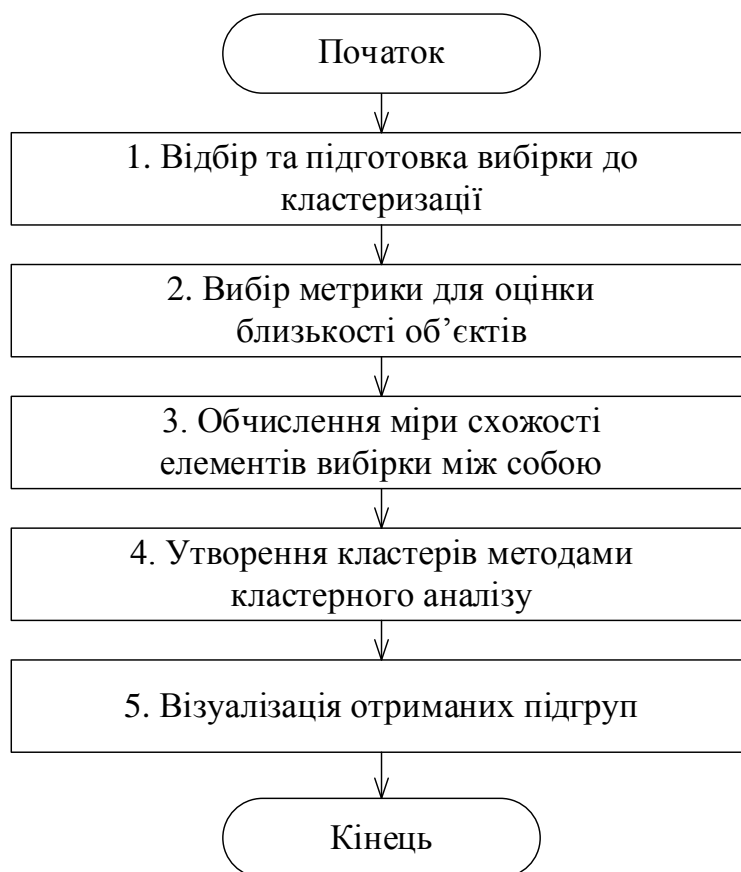


Рисунок 2.4 – Блок-схема загальної процедури кластеризації

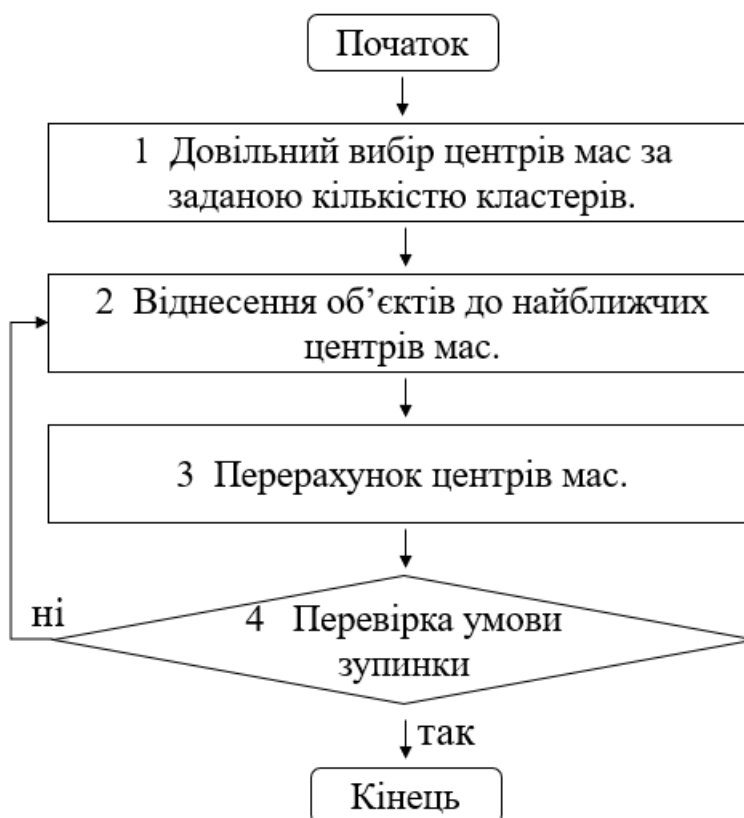


Рисунок 2.5 – Блок-схема кластеризації методом k-середніх

собою середній вигляд усіх характеристик форми наявних у кластері реалізацій [8, 27, 28, 31].

Тобто, як впливає з рисунку 2.4, на виході процедури, на відміну від навчання з учителем, де застосовується один еталон на один тип сигналів, тут матимемо певну кількість еталонів, яка дорівнює кількості кластерів.

Отже. Процедура формування еталонів у комбінованому навчанні здійснюється за тим же механізмом, що і при навчанні з учителем – еталонами є середній вигляд характеристик форми сигналів. Але на відміну від навчання з учителем, усереднення застосовується не в межах одного класу, а в межах кластерів цього класу. Це призведе до появи деякої кількості еталонів в кожному класі, що рівна кількості кластерів.

Рисунок 2.6 ілюструє такий випадок. Еталонами охоплюється набагато більший простір точок, ніж при використанні одного еталона, утвореного при навчанні з учителем.

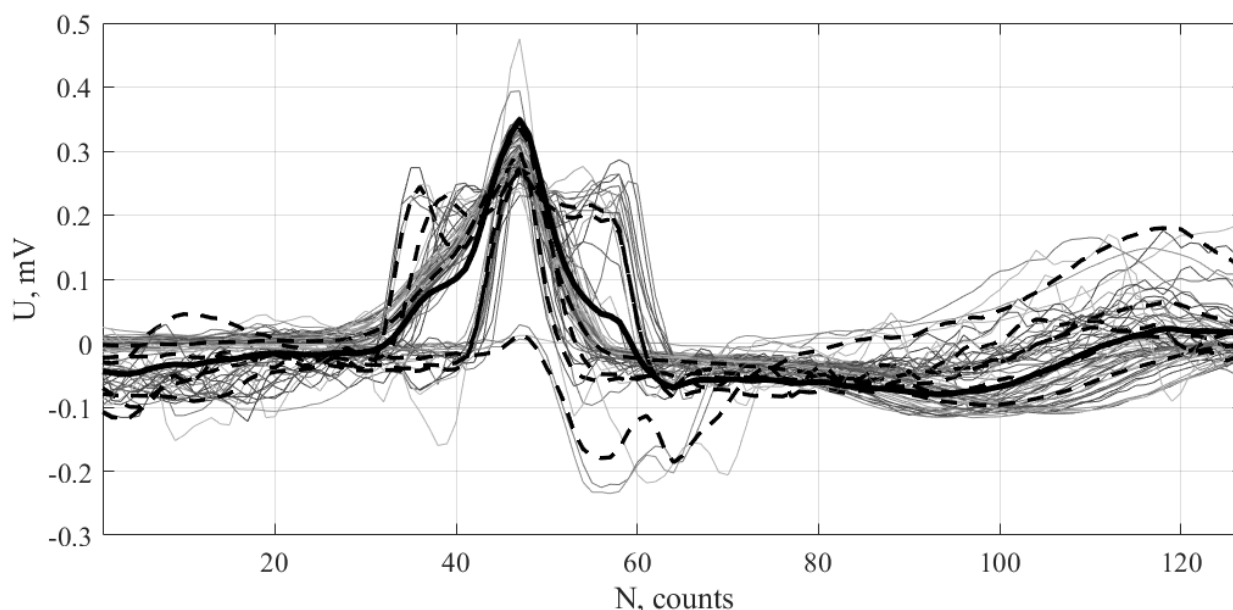


Рисунок 2.6 – Еталон, утворений методом навчання з учителем (суцільна лінія) та еталони, утворені процедурою комбінованого навчання (штрих-пунктир) на тлі реалізацій навчальної вибірки, на прикладі V-типу QRS-комплексів ЕКГ

З міркувань простоти та поставлених вимог до методу кластеризації

використовується стандартний алгоритм k -середніх. Кількість реалізацій та кластерів у вибірках однакова.

Лістинг робочої програми, що виконує задачу кластеризації навчальних вибірок та утворення еталонів, у випадку комбінованого навчання, наведено в додатку В.

Таким чином, прогнозується підвищення результативності розпізнавання медико-біологічних сигналів. Використання простого та швидкого алгоритму кластеризації не повинно сильно ускладнювати роботу алгоритму додатковими розрахунками.

Програмна реалізація завдання попередньої обробки даних та формування еталонів, у випадку навчання зі вчителем, наведена у додатку А.

2.4 Процедура прийняття рішення про тип вхідного сигналу

Як вже зазначалося, процедура прийняття рішень у разі комбінованого навчання алгоритму розпізнавання відбувається за тим же принципом, що і процедура прийняття рішень при навчанні з учителем (рис. 2.7), тобто шляхом порівняння вхідної реалізації з еталонами кожного з відомих типів і прийняття рішення на користь того типу, еталон якого дає найбільшу схожість.

Комбінування навчання призводить до появи більшої кількості еталонів, а тому процедура прийняття рішення дещо доповнюється (рис. 2.8). Під доповненням розуміється введення додаткового етапу, на якому визначається належність найбільш схожого еталона до заданого учителем типу. Класифікація вхідній реалізації присвоюється за типом найбільш схожого еталона.

Для оцінки схожості застосовуються коефіцієнти кореляції z – безрозмірні статистичні показники лінійної залежності двох величин одна від одної, значення яких лежать на проміжку $[-1,1]$, де -1 – вказує на повну їх відмінність (протилежність), а 1 – повний збіг послідовностей відліків величин.

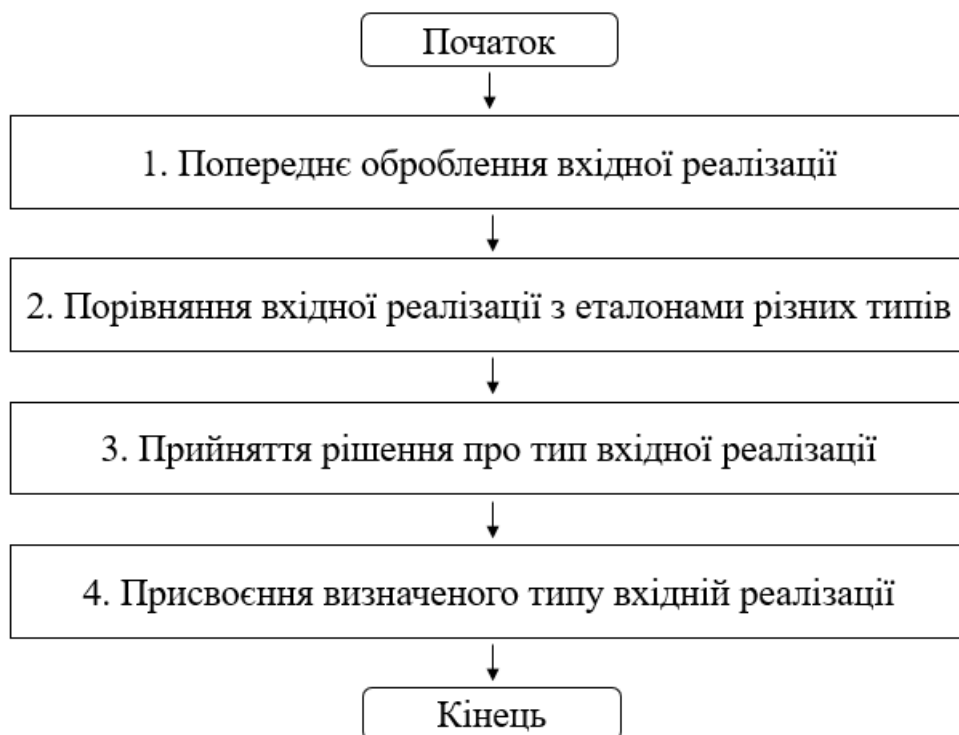


Рисунок 2.7 – Блок-схема прийняття рішень при навчанні з учителем

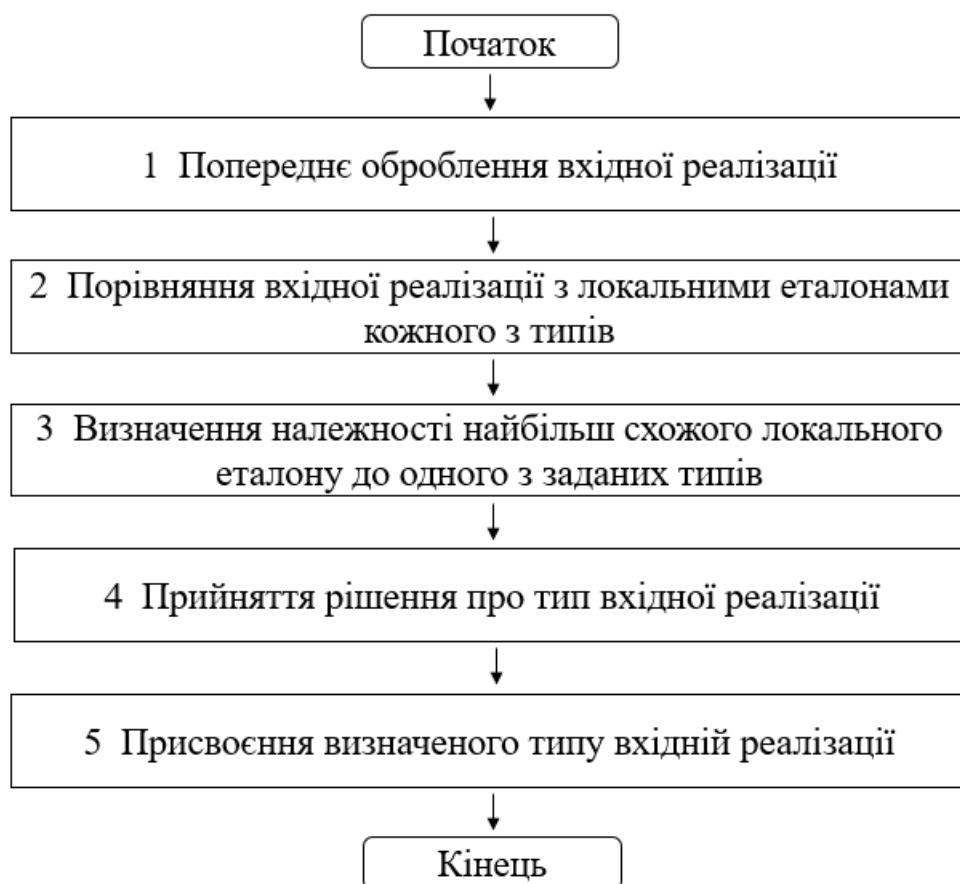


Рисунок 2.8 – Блок схема процедури прийняття рішень при комбінованому навчанні алгоритму розпізнавання

Тобто, чим ближчим до одиниці буде коефіцієнт кореляції – тим більша схожість форми вхідної реалізації з еталоном. Для зручності застосування коефіцієнт переводиться на шкалу $[0,1]$ за формулою (2.1).

$$z^* = \frac{1}{2}z + \frac{1}{2}, \quad (2.1)$$

де z – коефіцієнт кореляції.

В роботі, як коефіцієнт кореляції, кількісний показник близькості форми застосовується скалярний добуток сигналів з еталонами.

Отже. Механізм процедури прийняття рішень при комбінованому навчанні залишається аналогічним до базової і зводиться до таких положень.

1. Знаходження скалярного добутку вхідної реалізації з кожним еталоном та переведення його значення на шкалу $[0,1]$.
2. Визначення найбільшого значення скалярного добутку та відповідного еталону.
3. Визначення належності еталону до класу.
4. Присвоєння типу вхідній реалізації за типом найбільш схожого еталона.

Використання загальновідомого механізму прийняття рішень, що не потребує додаткових складних розрахунків, додає універсальності запропонованому методу навчання. Єдиним недоліком такого доповнення є можливе досить значне збільшення кількості розрахунків, до чого призводить збільшення кількості еталонів та, відповідно, порівнянь з ними [1].

Програмна процедура прийняття рішень наведена у додатках Б та В для навчання з учителем та комбінованого навчання відповідно.

2.5 Висновки до розділу 2

В даному розділі наводилася характеристика програмної реалізації та загального математичного змісту процедур, які є складовими в алгоритмах комбінованого навчання та розпізнавання сигналів за результатами такого навчання. Був розкритий зміст пропозицій щодо модифікації розпізнавального алгоритму та процедури обробки сигналів, що прогнозовано покращить роботу комбінованого алгоритму. Загальні висновки та результати розробок розділу 2 характеризуються такими положеннями.

1. Визначено дії учителя з підготовки та обробки даних для формування навчальних і контрольних вибірок.

2. Розглянуто процедури попередньої обробки даних – приведення їх до характеристики форми, з математичної та програмної сторони. Розроблено відповідні алгоритми.

3. Розглянуто поняття вибірки, процедури утворення навчальних та контрольних вибірок з генеральної вибірки даних. Уточнено умови формування вибірок. Обрано процедуру послідовної вибірки даних, тобто обрання сигналів з наявної вибірки в запису ЕКГ з певним інтервалом, заданим вручну.

4. Уточнено процедуру кластеризації, метод, спосіб застосування та можливі варіанти керування роботою процедури. Визначено спосіб модифікації цієї процедури з використанням кореляційної метрики – приведенням у відповідність критерія відбору сигналів в кластери до критерію прийняття рішень про типи вхідних реалізацій сигналу під час їх розпізнавання.

Сформовано програмну процедуру утворення еталонів у разі комбінованого навчання та для процедури навчання з учителем.

5. Розроблено процедури прийняття рішень для комбінованого навчання у вигляді доповненої процедури прийняття рішень для алгоритму навчання з учителем.

6. Узагальнено деякі особливості та властивості розроблених процедур, які повинні братися до уваги під час їх застосування та проведення тестової перевірки результативності процедур комбінування машинного навчання, про які йтиметься у наступному розділі.

7. Лістинг робочих програм, реалізованих у середовищі MATLAB, наведено у додатках А, Б та В.

3 ЧИСЕЛЬНИЙ ЕКСПЕРИМЕНТ З ОЦІНКИ РЕЗУЛЬТАТИВНОСТІ КОМБІНОВАНОГО НАВЧАННЯ АЛГОРИТМІВ РОЗПІЗНАВАННЯ МЕДИКО- БІОЛОГІЧНИХ СИГНАЛІВ

3.1 Мета і завдання чисельного експерименту

Основною метою чисельного експерименту є перевірка прогнозованих змін [33] у результативності розпізнавання сигналів із використанням запропонованого методу комбінованого навчання алгоритму розпізнавання медико-біологічних сигналів.

Під час розгляду теоретичної частини роботи зазначалося, що введення додаткової кластеризації в межах заданих учителем типів, повинно сприяти більш детальному урахуванню особливостей та часткових випадків форми, що характерні для досліджуваного типу медико-біологічних сигналів.

Тестова перевірка (рис. 3.1) в цій роботі буде проведена для алгоритму навчання з учителем, розробленого алгоритму його комбінованого навчання з різними метриками кластеризації.

Змістом перевірки є підрахунок загальної кількості прийнятих досліджуваним алгоритмом на контрольних вибірках рішень, кількості вірних рішень та розрахунок значень критеріїв чутливості і специфічності алгоритмів. За порівнянням критеріїв специфічності буде сформовано висновки про доцільність введення алгоритму комбінованого навчання.

В цілому, чисельний експеримент планувався як тестова перевірка розроблених алгоритмів на прикладі роботи з реальними медико-біологічними сигналами пацієнта в їх запису в базі даних в Інтернет.

В цілому, даний підрозділ розкриває мету та зміст тестової перевірки алгоритмів комбінованого навчання, що описується такими положеннями.

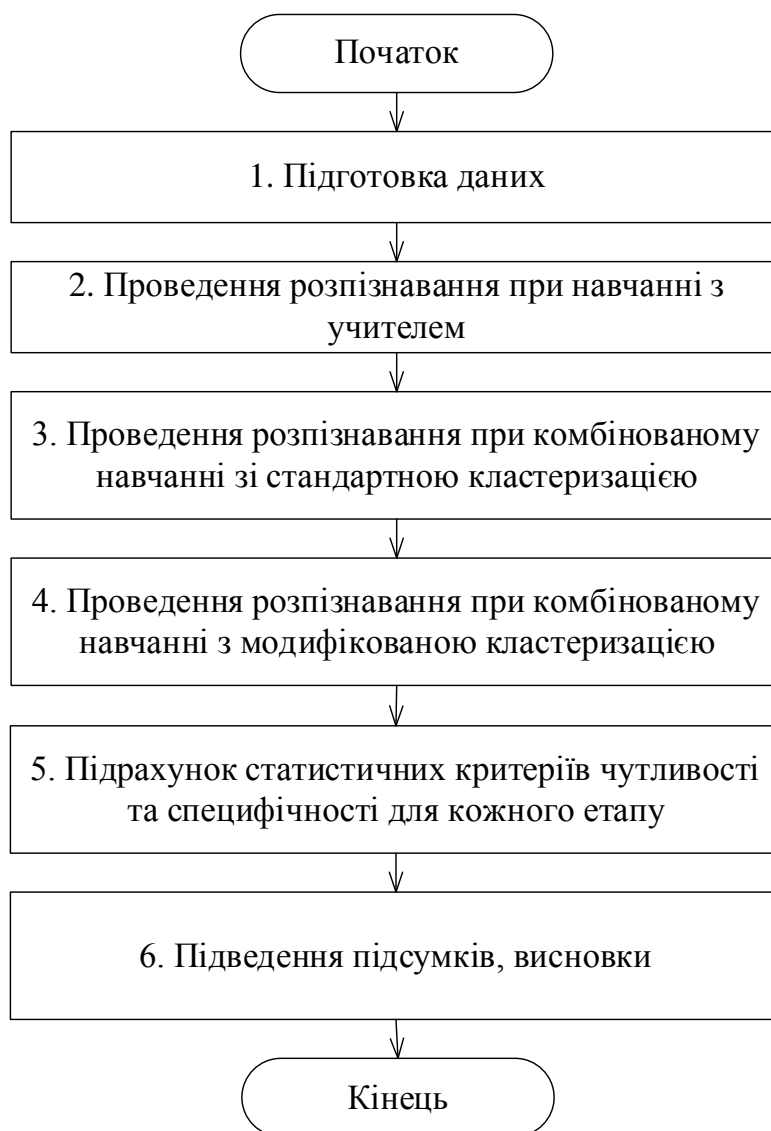


Рисунок 3.1 – Блок-схема проведення чисельного експерименту

В цілому, даний підрозділ розкриває мету та зміст тестової перевірки алгоритмів комбінованого навчання, що описується такими положеннями.

1. Обрані цілі направлені на перевірку спроможності виконаних досліджень та розробок щодо комбінованого навчання розпізнавальних алгоритмів, представлених в роботі.

2. Оцінка результативності роботи алгоритму – за критеріями чутливості та специфічності.

3. Розглядалася перевірка процедур навчання із використанням базового та розробленого модифікованого до комбінованого навчання алгоритмів із подальшим порівнянням результатів.

3. Складена блок-схема чисельного експерименту.

4. Задача розпізнавання розглядається на реальному прикладі медико-біологічних сигналів пацієнта.

Більш детально про використанні сигнали та підготовку даних до проведення чисельного експерименту йдеться у наступному пункті.

3.2 Критерії статистичної оцінки результативності навчання алгоритму розпізнавання сигналів

В роботі для оцінки результативності роботи навчального алгоритму використовувалися статистичні критерії чутливості та специфічності алгоритму розпізнавання сигналів після його навчання, які знаходилися для кожного з дослідів, із подальшим порівнянням.

Критерії чутливості та специфічності є базовими статистичними показниками, що застосовуються у медичній діагностиці та є також традиційними для біології [4].

Чутливістю вважається параметр, що відображає долю правильно класифікованих результатів від загальної кількості об'єктів класифікації. Розраховується як відношення кількості вірно класифікованих реалізацій до загального обсягу вибірки окремого типу.

Під специфічністю розуміється доля вірно класифікованих результатів серед усієї кількості результатів (в тому числі і помилкових), прийнятих на користь окремого типу сигналів. Розраховується співвідношенням кількості вірно прийнятих рішень на користь окремого типу до загальної кількості рішень прийнятих.

Величини статистичних критеріїв у позитивному випадку повинні прямувати до 100%. Зауважимо, що на практиці, спроби підвищення чутливості часто

призводять до зниження специфічності [8, 9], так як це часто супроводжуються збільшенням числа хибно-позитивних результатів (рішень про типи сигналів).

Отже. Статистичними критеріями для оцінки результативності роботи алгоритму обрано чутливість та специфічність, що є базовими статистичними показниками, застосовуваними у медичній діагностиці.

Для обрахунку критеріїв програмною процедурою передбачається підрахунок кількості вірно та помилково прийнятих рішень та використання стандартних формул.

Оцінка специфічності та чутливості відбувається при умові розпізнавання системою сигналів контрольних вибірок.

3.3 Дані для проведення чисельного експерименту

Для перевірки розроблених алгоритмів у реальних умовах застосовувалися півгодинні записи електрокардіограми пацієнта, які зареєстровані у стандартних 12 відведеннях при частоті дискретизації 257 Гц. Запис ЕКГ (рис. 3.2) містить

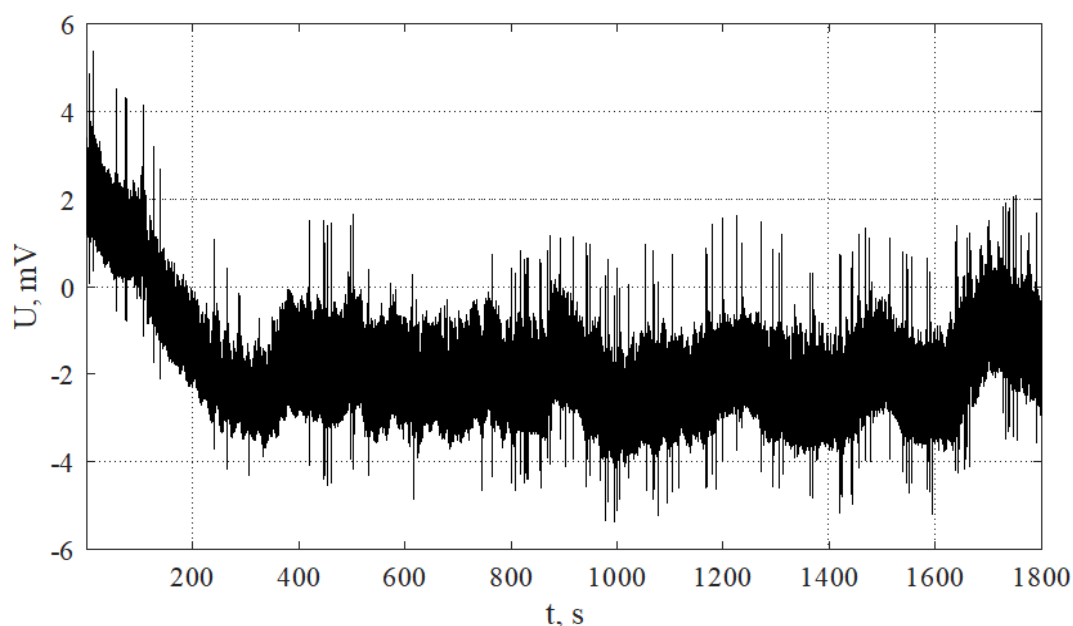


Рисунок 3.2 – Загальний вигляд використаного півгодинного запису електрокардіограми пацієнта

коментарі про тип кожного QRS-комплексу в запису та координати R-піку.

Для утворення вибірки QRS-комплексів було створено програмну процедуру, що фрагментує сигнал на QRS-комплекси на проміжку заданої кількості відліків кожного з них, що називається в роботі вікном та розподіляє ці комплекси по групах за коментарями про тип комплексу в цьому запису (N – Normal beats, A – Atrial beats, V – Premature ventricular contraction).

За відомими координатами R-піків під час відбору для забезпечення можливості їх порівняння зразки таких сигналів синхронізуються у вікні спостереження загальним обсягом у 128 відліків із положенням R-піку кожного сигналу на 48 відліку у цьому вікні. Ліворуч від положення R-піку рівномірно розташовано 47 відліків сигналу, праворуч – ще 80. Всього вікном спостереження охоплено інтервал часу тривалістю приблизно 500 мс.

Програмною процедурою відбору вибірок передбачено реалізацію відбору комплексів відповідно до їх кількості в запису. Таким чином, виконується умова формування вибірок – рівномірний відбір зразків по всій протяжності досліджуваного сигналу.

Дані про загальну кількість зразків комплексів в запису та значення кроку відбору сигналів з всієї вибірки зведено до таблиці 3.1.

Таблиця 3.1 – Кількість та довжина кроків по запису під час вибірки зразків QRS-комплексів по їх типах

Тип сигналів	N	A	V
Кількість зразків комплексів в запису ЕКГ	2816	<u>124</u>	136
Крок відбору комплексів з запису ЕКГ	22	1	1

Кількість сигналів у вибірках обмежена мінімальним обсягом вибірки по всім типами. У досліджуваному випадку найменший обсяг вибірки спостерігається по вибірці А-типу – 124 QRS-комплекси, саме тому кількість реалізацій для N та V типів також обмежена цим числом.

Відібрана вибірка програмно поділяється на навчальну та контрольну з рівномірним кроком в один комплекс, тобто по чергово. Обсяг кожної навчальної та контрольної вибірки – 62 QRS-комплекси.

Отже. Підрозділ розкриває основний зміст даних для тестової перевірки розроблених процедур навчання і розпізнавання та алгоритми утворення вибірок, що характеризується наступними положеннями.

1. Для тестової перевірки застосовано реальний сигнал ЕКГ пацієнта, зареєстрований на частоті дискретизації 257 Гц у 12 відведеннях з анотацією про типи QRS-комплексів та координатами R-піків. У дослідження використовуються дані одного відведення.

2. Відібрані QRS-комплекси синхронізуються у вікні 128 відліків за R-піком на 48 відліку для забезпечення порівняння форми алгоритмами їх розпізнавання і проведення досліджень в цілому.

3. Виконується умова формування вибірок – рівномірний відбір зразків із загальної сукупності в запису ЕКГ. Відбір здійснюється з автоматичним розрахунком інтервалу рівномірного відбору сигналів, виходячи з загального обсягу вибірки, загальної їх кількості в запису.

4. Кількість відібраних сигналів передбачено за найменшим обсягом з вибірок комплексів по їх типам. Відібрано 124 реалізації комплексів в кожному класі, які потім розподілені на навчальну і контрольну вибірку з інтервалом в один комплекс по 62 реалізації на кожну.

5. Програмно реалізовано вищеописані процедури в середовищі MATLAB.

Зміст програмної реалізації алгоритмів у середовищі MATLAB розкривається у наступному підрозділі.

3.4 Програмні процедури середовища MATLAB для проведення чисельного експерименту

Як вже згадувалося в попередньому підрозділі, розробка програмних процедур та розрахунки відбувалася у середовищі MATLAB. Вибір саме цього програмного інструментарію обумовлюється тим, що в даному середовищі наявні всі необхідні інструменти та засоби роботи із масивами значних обсягів, тобто для обробки та аналізу сигналів, які потрібні в роботі. Зокрема, в цьому середовищі наявна значна кількість вбудованих функцій кластеризації сигналів, що спрощує написання алгоритмів машинного навчання. Окрім цього, MATLAB має потужне графічне середовище [18, 19], що дозволяє візуалізувати необхідні графічні дані у повному потрібному в роботі обсязі.

Поетапне проведення експерименту описується такими функціями середовища Matlab.

1. Підготовка та попередня обробка вибірок даних:

- *fopen, fread, textread* – відкриття та читання файлів вибірки з Internet бази даних;
- *plot* – візуалізація вхідного сигналу (рис. 3.2) та синхронізованих QRS-комплексів (рис. 3.3);
- *medfilt1* – медіанний фільтр для стабілізації ізолінії сигналу, подальший розрахунок характеристики форми базується на математичних виразах та формулах.

2. Проведення процедури навчання з учителем – базується на математичних виразах та формулах, зокрема, з використанням математичних операторів *dot* та *mean* [18, 19] – для знаходження скалярного добутку сигналів та середнього значення відповідно.

3. Процедура комбінованого навчання – аналогічно до навчання з учителем базується на формулах та використанні аналогічних математичних операторів, зокрема:

- *kmeans* – реалізує кластеризацію методом k-середніх із застосування відстані Евкліда за замовчуванням;
- *'distance', 'correlation'* – синтаксис кореляційної відстані (критерій α_z)

Подальші етапи підрахунку статистичних критеріїв та кількості прийнятих рішень відбуваються за використання формул та математичних виразів і не застосовують додаткових функцій середовища, зокрема, функцій графічної візуалізації результатів *plot* та *stem* [18, 19], що реалізують графіки та стовпчикові діаграми прийнятої кількості рішень, відповідно.

Отже, результатом розгляду даного підрозділу є характеристика використаних функцій середовища MATLAB та опис їх застосування на етапах проведення чисельного експерименту тестової перевірки розроблених алгоритмів комбінованого машинного навчання.

3.5 Результати чисельного експерименту та їх аналіз

Першим етапом експериментальних досліджень (рис. 3.1) після процедури підготовки даних до розпізнавання є процедура навчання з учителем.

Як зазначалося, під процедурою навчання з учителем розуміється отримання еталонів, що є середнім виглядом вхідних реалізацій даного типу. У такому випадку на кожен з заданих учителем типів припадає один еталон, що утворений як середнє значення характеристик форми всіх реалізацій навчальної вибірки даного типу.

Для навчання в розрахунках застосовано 62 реалізації навчальної вибірки для кожного з типів сигналів, заданих учителем, що пройшовши попередню обробку подавалися на вхід алгоритму. Так як кількість заданих типів у випадку QRS-комплексів сигналів електрокардіограми рівна 3, то відповідно на виході алгоритму маємо 3 еталонних сигнали (рис. 3.3).

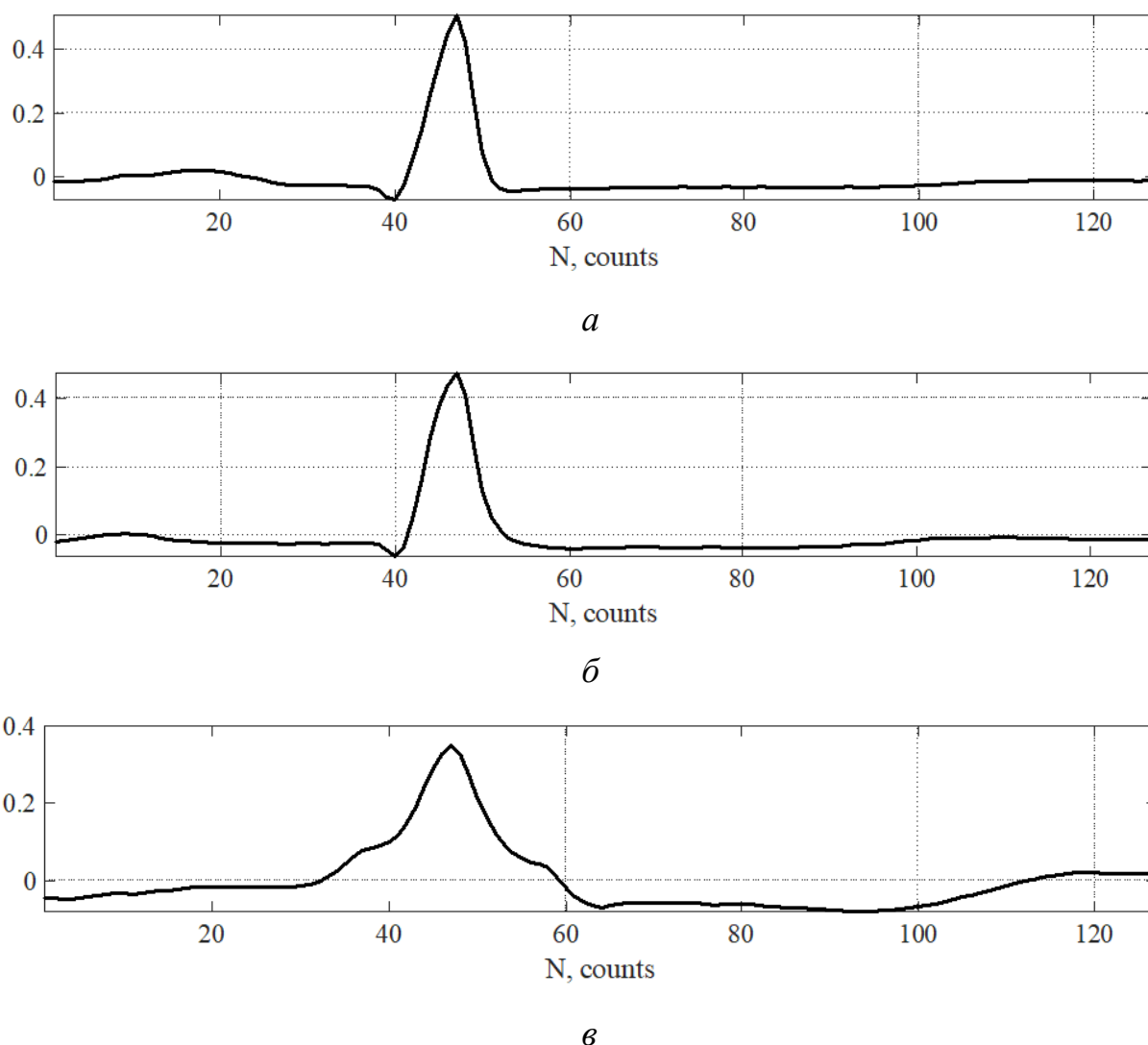


Рисунок 3.3 – Еталони, утворені в результаті навчання з учителем:

a – N, *б* – A, *в* – V- тип

Після навчання проводиться процедура розпізнавання сигналів контрольної вибірки. Відповідно до підрозділу 3.3, обсяг контрольної вибірки дорівнює обсягу навчальної вибірки, тобто контрольна вибірка складається із трьох груп по 62 сигналу у кожній.

Результати та значення статистичних критерії якості розпізнавання QRS-комплексів електрокардіограми зведено до таблиці 3.2.

Другим етапом експерименту було заплановано перевірку розробленого алгоритму комбінованого навчання.

Таблиця 3.2 — Результати та значення статистичних критерії якості розпізнавання QRS-комплексів контрольної вибірки алгоритмом, навченим з учителем, з одним еталоном на клас сигналів

Типи QRS-комплексів

Типи рішень	N	A	V	Специфічність, %
N	<u>57</u>	14	2	78,08
A	5	<u>48</u>	2	87,27
V	0	0	<u>58</u>	100
Чутливість, %	91,94	77,42	93,55	<u>89,25</u>

Експеримент проводитиметься у два кроки, на першому для кластеризації буде застосована відстань Евкліда, на другому – кореляційна відстань за критерієм α_z (формула (1.3)) із подальшим порівнянням результатів. Таке порівняння дозволить визначити вплив обраної метрики та узгодження між критеріями прийняття рішень та відбору еталонів на якість кластеризації та розпізнавання в цілому.

Метод кластеризації в алгоритмі – k-середніх. Кількість кластерів – 6.

Число кластерів підбрано так, щоб на кожний кластер в середньому приходилось по 10 реалізацій.

Кількість реалізацій у кластерах є нерівномірною, в залежності від особливостей форми сигналів (в часткових випадках кластери можуть мати один екземпляр сигналу).

Результат проведеної в дослідженнях кластеризації [8], тобто розподіл реалізацій контрольної вибірки по кластерам зображено на діаграмі (рис. 3.4) та зведено до таблиці (табл. 3.3).

На цих діаграмах наведено розподіл елементів по кожному з шести кластерів. Зокрема, в шостому кластері сигналів А типу 7 екземплярів з вибірки з №№ 10, 16, 28, 36, 47, 56, 57, в першому кластері таких екземплярів – 18 і так далі.

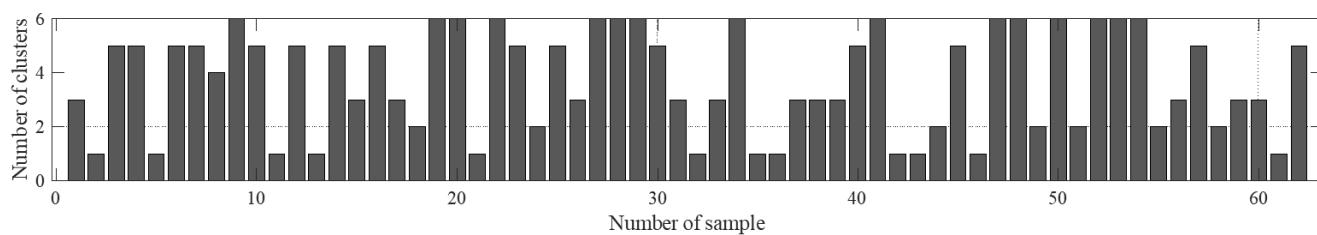
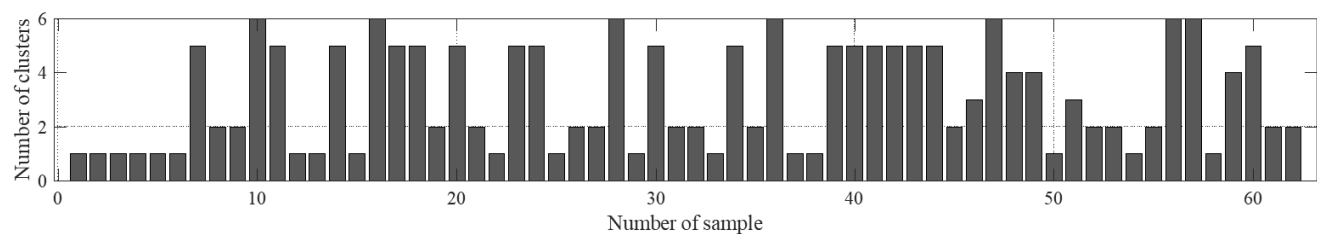
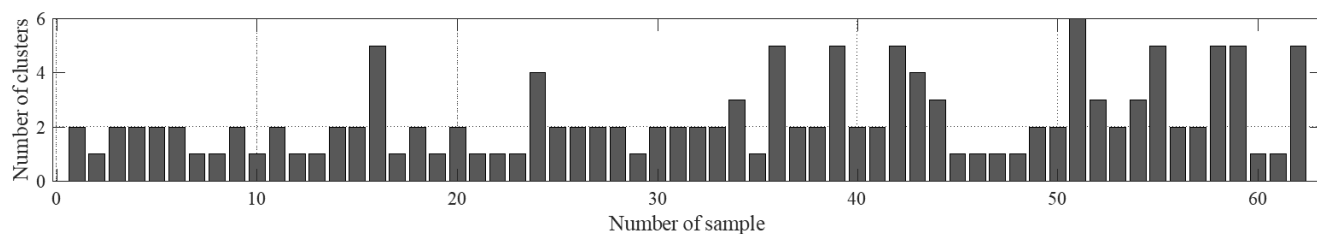
*a**б**в*

Рисунок 3.4 – Розподіл елементів вибірки по кластерах при застосуванні
Евклідової метрики: *a* – N, *б* – A, *в* – V- тип

Таблиця 3.3 — Кількість реалізацій сигналу у кластері при застосуванні
Евклідової метрики

№ кластера	1	2	3	4	5	6
N-тип						
Кількість сигналів	12	7	12	1	15	15
A-тип						
Кількість сигналів	18	15	2	3	17	7
V-тип						
Кількість сигналів	19	28	4	2	8	1

Особливістю реалізації даного методу кластеризації в середовищі MATLAB є неможливість зафіксувати номер кластеру [18, 19] – з кожним запуском нумерація кластерів змінюється. Для запобігання цього в алгоритмі передбачено збереження результатів кластеризації та їх завантаження із новим запуском. Еталонні значення отримуються усередненням всіх реалізацій в межах кожного кластеру. В результаті усереднення сигналів в кластерах утворюється сімейство еталонів, кількість яких для кожного типу рівна кількості кластерів вибірки цього типу (рис. 3.5).

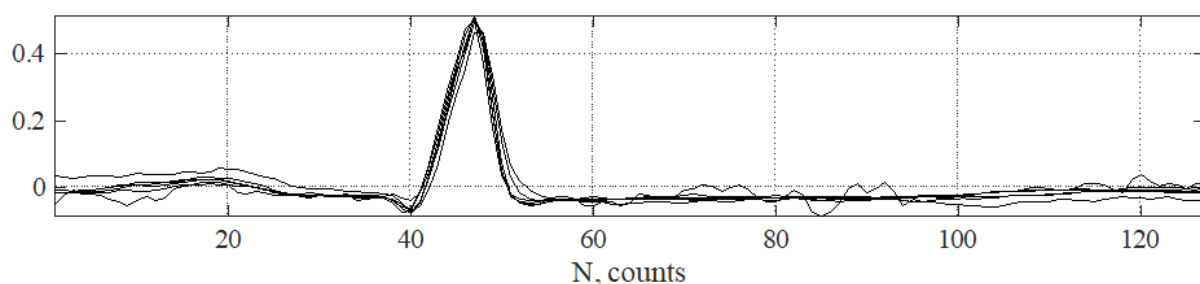
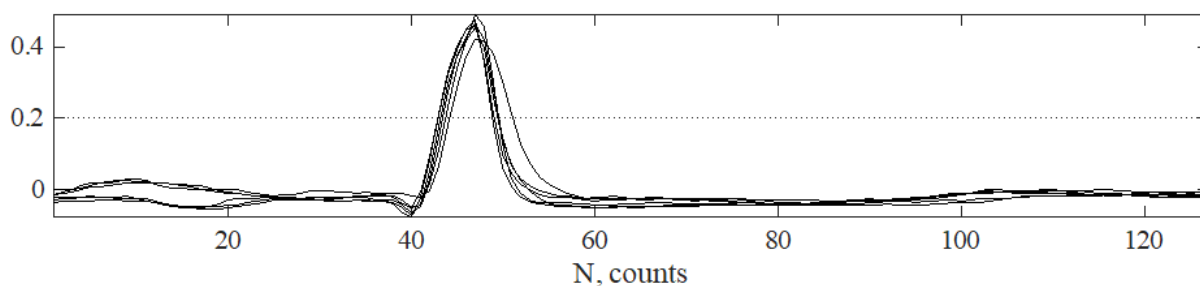
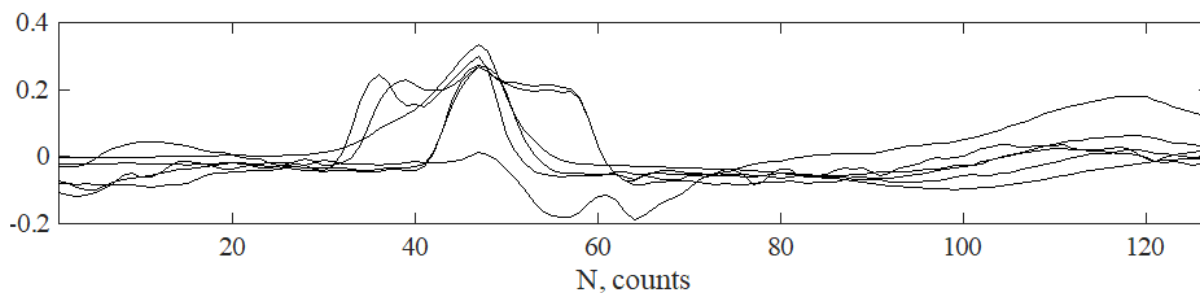
*a**б**в*

Рисунок 3.5 – Еталони, утворені в результаті комбінованого навчання із застосуванням Евклідової метрики: *a* – N-тип, *б* – А-тип, *в* – V- тип

Контрольна вибірка аналогічна застосованій у попередньому досліді, що дозволить порівняти результативність роботи алгоритму (табл. 3.4).

Таблиця 3.4 — Результати та статистичні критерії якості розпізнавання контрольної вибірки при комбінованому навчанні з використанням Евклідової метрики

Типи рішень	Типи QRS-комплексів			Специфічність, %
	N	A	V	
N	<u>58</u>	6	0	90,63
A	4	<u>56</u>	2	90,32
V	0	0	<u>60</u>	100
Чутливість, %	93,65	90,32	96,77	<u>93,55</u>

Розпізнавання у випадку комбінованого навчання проводиться шляхом порівняння вхідної реалізації з кожним із еталонів, після чого визначається належність найбільш схожого з них до заданого учителем типу, за яким відбувається присвоєння результату класифікації вхідній реалізації.

Наступний крок – аналогічний дослід із використанням комбінованого навчання, відмінність якого полягає у використанні під час кластеризації вибірок критерію кореляційної відстані α_z (формула (1.3)).

Таке доповнення повинне узгодити процедуру кластеризації із алгоритмом прийняття рішень з використанням коефіцієнтів кореляції – скалярних добутків сигналів з еталонами.

Метод кластеризації та кількість кластерів залишається незмінним. Результати кластеризації можуть суттєво залежати від обраної метрики. Розподіл реалізацій навчальної вибірки по кластерам, отриманий під час досліджень, зведено до діаграми (рис. 3.6) та таблиці (табл. 3.5).

Як і в попередньому досліді, в результаті кластеризації утворюється сімейство еталонів, кількість яких рівна кількості кластерів, кожен з яких є середнім виглядом усіх реалізацій в межах свого кластеру (рис. 3.7).

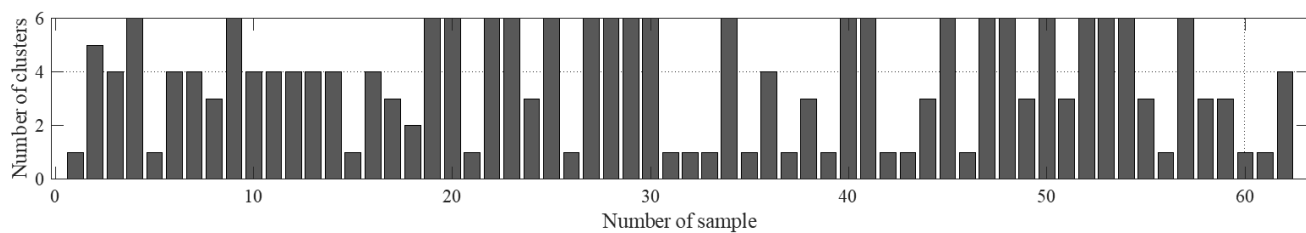
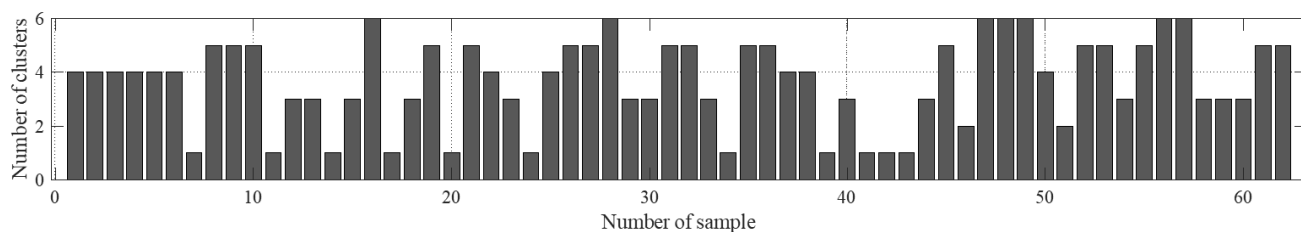
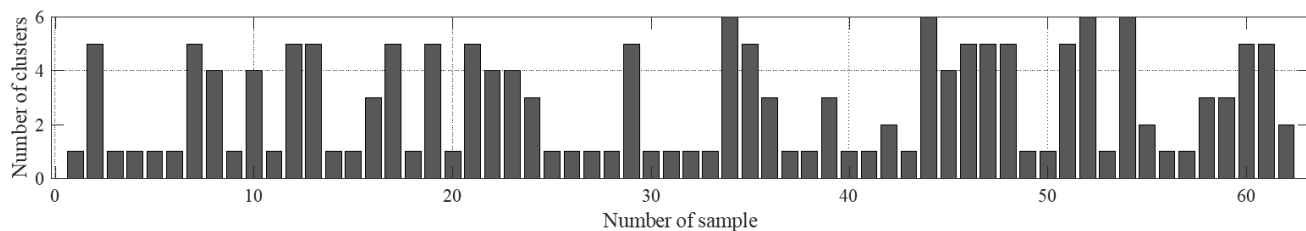
*a**б**в*

Рисунок 3.6 – Розподіл елементів вибірки по кластерах при застосуванні критерію

 α_z : *a* – N-тип, *б* – A-тип, *в* – V- типТаблиця 3.5 — Кількість реалізацій сигналу у кластері при застосуванні критерію α_z

№ кластера	1	2	3	4	5	6
N-тип						
Кількість сигналів	17	1	10	11	1	22
A-тип						
Кількість сигналів	11	2	14	11	17	7
V-тип						
Кількість сигналів	29	3	6	5	15	4

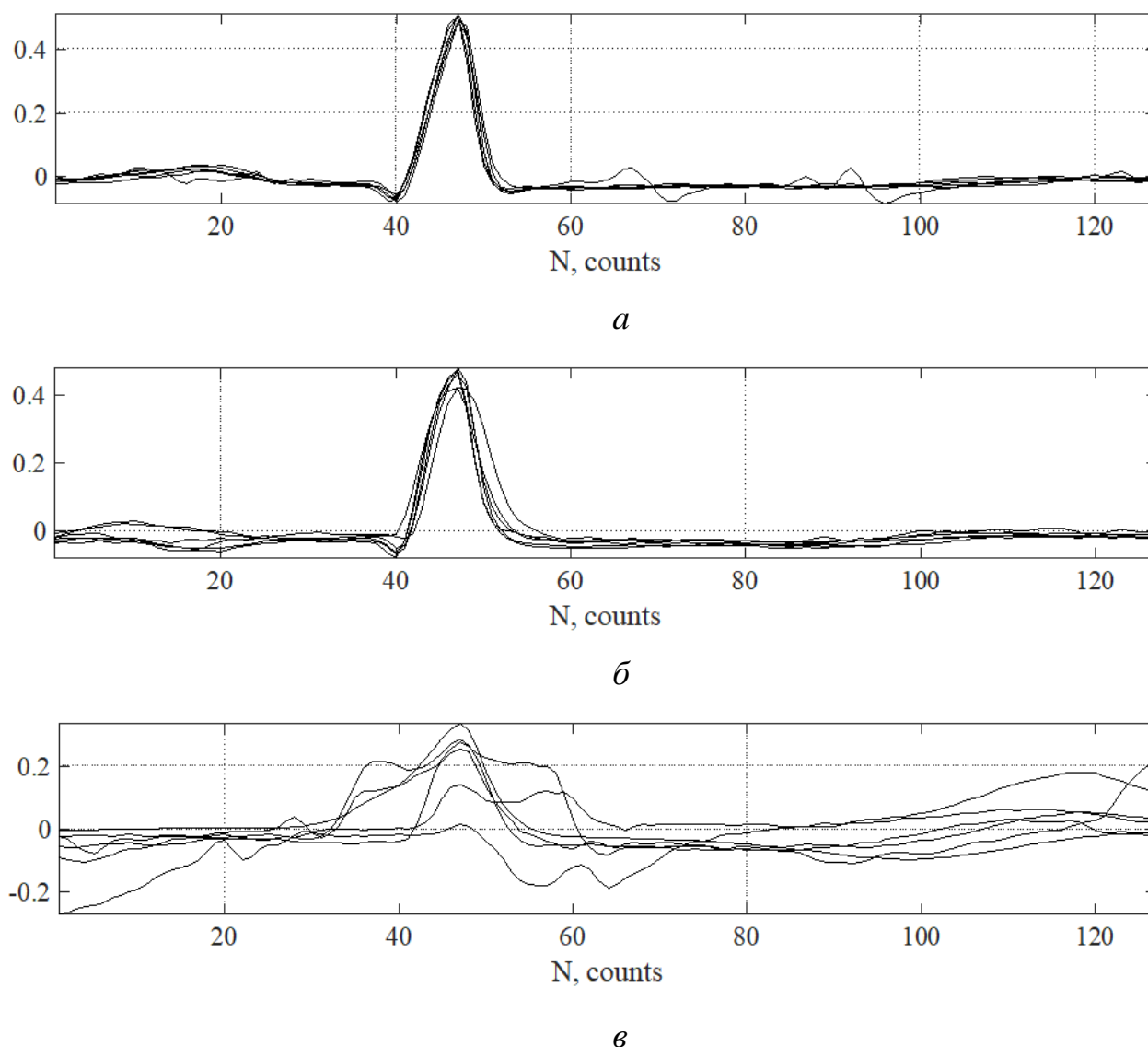


Рисунок 3.7 – Еталони, утворені по типах сигналів в кластерах в результаті комбінованого навчання із застосуванням критерію α_z : *a* – N-тип, *б* – А-тип, *в* – V-тип

Отримані результати розпізнавання контрольної вибірки алгоритмом, навченим із застосуванням кластеризації з метрикою α_z візуалізовано у таблиці 3.6.

На даному етапі основні робота з дослідження основних алгоритмів комбінованого навчання завершується.

На заключному етапі експерименту відбувалося порівняння та аналіз отриманих для різних варіантів навчання результатів, за якими буде підведено загальні підсумки експерименту та роботи в цілому.

Таблиця 3.7 — Результати та статистичні критерії якості розпізнавання контрольної вибірки при комбінованому навчанні із застосуванням критерію α_z

Типи QRS-комплексів

Типи рішень	N	A	V	Специфічність, %
N	<u>59</u>	5	1	90,77
A	3	<u>57</u>	1	93,44
V	0	0	<u>60</u>	100
Чутливість, %	95,16	91,94	96,77	<u>94,62</u>

Для більш чіткого розуміння зведемо результати експерименту до загальних стандартних таблиць 3.8 – 3.13 та подамо їх також у форматі діаграм (рис. 3.8 – 3.11).

Таблиця 3.8 — Чисельні значення кількості вірно прийнятих рішень при різних варіантах навчання розпізнавального алгоритму

Типи сигналів	N	A	V
Без кластеризації, %	57	48	<u>58</u>
k-середніх станд., %	58	56	<u>60</u>
k-середніх α_z , %	59	57	<u>60</u>

Таблиця 3.9 — Чисельні значення чутливості прийняття рішень при різних варіантах навчання розпізнавального алгоритму

Типи сигналів	N	A	V
Без кластеризації, %	91,94	77,42	93,55
k-середніх станд., %	93,65	90,32	96,77
k-середніх α_z , %	95,16	91,94	96,77

Таблиця 3.10 — Чисельні значення специфічності прийняття рішень при різних варіантах навчання розпізнавального алгоритму

Типи сигналів	N	A	V
Без кластеризації, %	78,08	87,27	100
k-середніх станд., %	90,63	90,32	100
k-середніх α_z , %	90,77	93,44	100

Таблиця 3.11 — Зміна кількості вірно прийнятих рішень у порівнянні рішеннями базового розпізнавального алгоритму з одним еталоном в кожному класі сигналів при кластеризації з різними метриками

Тип сигналів	N	A	V
k-середніх станд.	+1	+8	+2
k-середніх α_z ,	+1	+1	-

Таблиця 3.12— Зміна відсотку чутливості рішень у порівнянні цим показником базового розпізнавального алгоритму з одним еталоном в кожному класі сигналів при кластеризації з різними метриками

Типи сигналів	N	A	V	Середнє
k-середніх станд., %	+1,71	+12,9	+3,22	+5,94
k-середніх α_z , %	+3,22	+14,52	+3,22	+6,99

Таблиця 3.13 — Зміна відсотку специфічності рішень у порівнянні цим показником базового розпізнавального алгоритму з одним еталоном в кожному класі сигналів при кластеризації з різними метриками

Типи сигналів	N	A	V	Середнє
k-середніх станд., %	+12,55	+3,05	-	+5,2
k-середніх α_z , %	+12,69	+6,17	-	+6,29

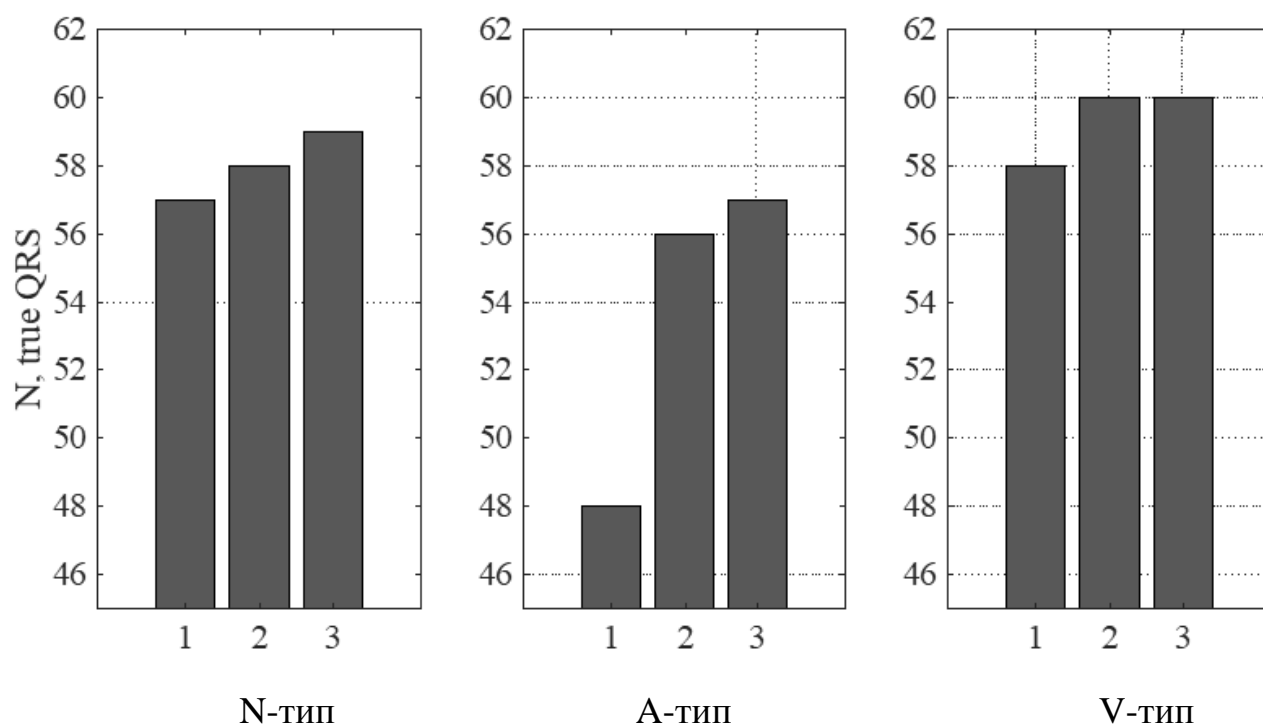


Рисунок 3.8 – Діаграми зміни правильних рішень по типам сигналів, прийнятих при звичайному навчанні із вчителем (1), комбінованому навчанні з Евклідовою метрикою (2) та з використанням метрики α_z (3)

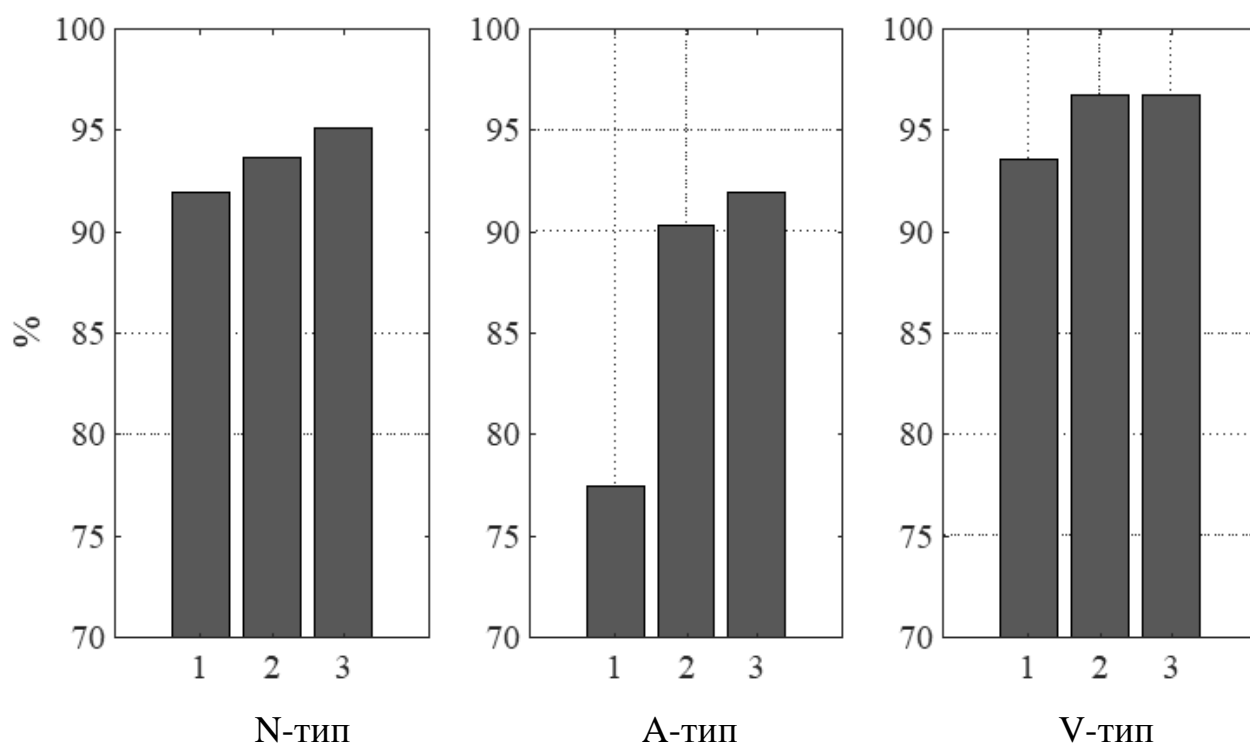


Рисунок 3.9 – Діаграми зміни показника чутливості по типам сигналів, прийнятих при звичайному навчанні із вчителем (1), комбінованому навчанні з Евклідовою метрикою (2) та з використанням метрики α_z (3)

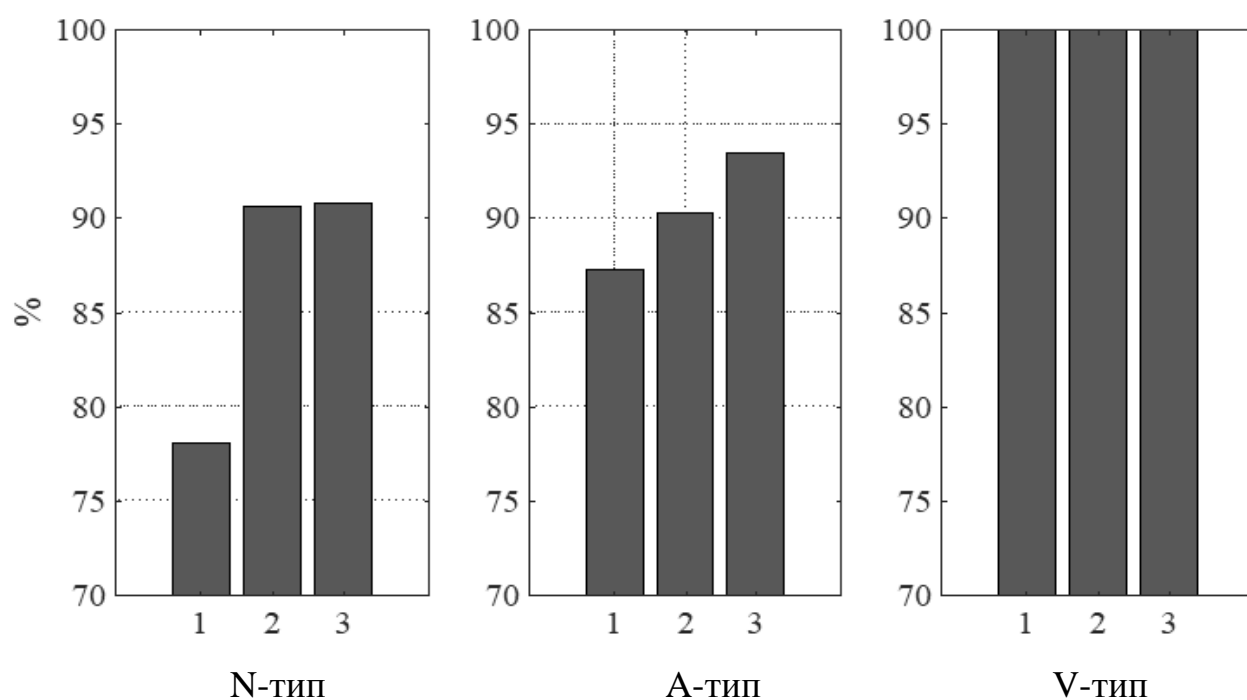


Рисунок 3.10 – Діаграми зміни показника специфічності по типам сигналів, прийнятих при звичайному навчанні із вчителем (1), комбінованому навчанні з Евклідовою метрикою (2) та з використанням метрики α_z (3)

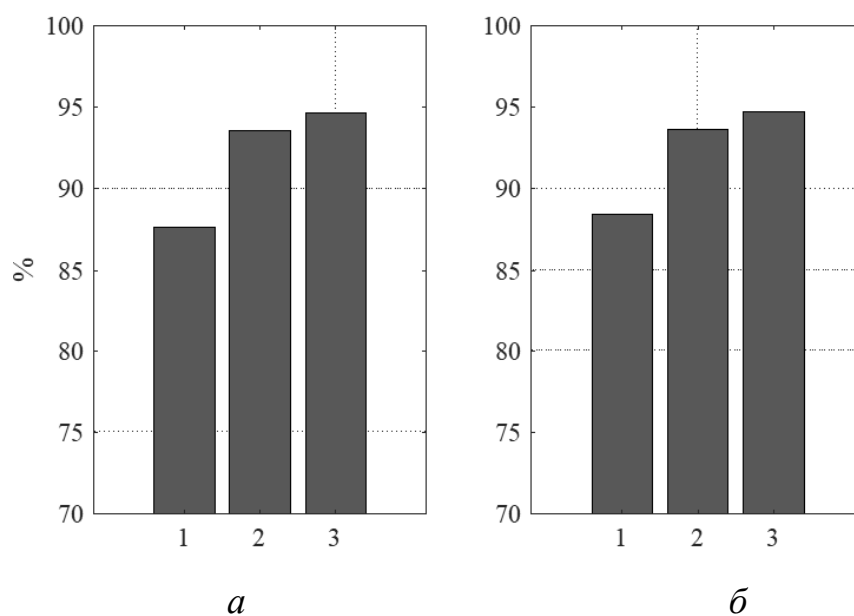


Рисунок 3.11 – Діаграми зміни середніх значень показників чутливості рішень (а) та специфічності рішень (б) по типам сигналів, прийнятих при звичайному навчанні із вчителем (1), комбінованому навчанні з Евклідовою метрикою (2) та з використанням метрики α_z (3)

Отже, головні результати дослідження в чисельному експерименті на тестовому прикладі було проілюстровано у діаграмах та таблицях. Сформовані діаграми вказують на те, що найбільшою кількістю вірно прийнятих рішень характеризується алгоритм, комбіновано навчений із застосуванням критерію α_z . Показники чутливості та специфічності також досягли найбільших значень саме при цьому типі кластеризації.

Мабуть, як скоріш за все здається, декілька більш високий рівень валідності рішень в тестовому прикладі з метрикою α_z під час навчання пояснюється тим, що у разі кластеризації навчальних вибірок з такою метрикою для формування системи еталонів (системи зразків сигналів по класах для проведення подальшого розпізнавання) та в самих алгоритмах прийняття рішень про типи вхідних сигналів, що таким чином навчалися, використовується той же самий скалярний добуток векторів, які порівнюються, тобто між цими критеріями у навчанні і в прийнятті рішень має місце повна відповідність.

Найнижчим показником специфічності рішень характеризувався N-тип, так як на його користь, при навчанні з учителем, помилково приймалися деякі сигнали A-типу. Відповідно A-тип мав найнижчий показник чутливості, так як значна частина його реалізацій класифікувалася як N-тип. Такий ефект виникає у зв'язку з високою подібністю сигналів даних типів.

Найвищі показники чутливості та специфічності, здобуті V-типом, так як його характеристика форми у більшості випадків не співпадає з іншими типами, та характеризується певним розкидом значень.

Перехід до комбінованого навчання позитивно вплинув на якість класифікації типів сигналів, форма сигналів яких збігається. Високий розрив між показниками чутливості та специфічності для типів таких сигналів скорочено.

Аналіз результатів експерименту вказує на спроможність виконаних досліджень і розробок. Розроблений програмний інструментарій забезпечує можливість перевірки результативності запропонованого алгоритму комбінованого навчання та оцінки резервів щодо підвищення вірності

розпізнавання типів сигналів шляхом введення додаткової їх кластеризації в межах кожного з типів.

Узгодження між метрикою кластеризації в процедурах навчання та в розпізнавальному алгоритмі в розглянутому тестовому прикладі сприяє підвищенню вірності розпізнавання типів сигналів.

Загалом, прогнозовані результати повністю підтвердились, проведені розробки можна вважати спроможними.

3.6 Висновки до розділу 3

Отже. Основні результати, здобуті під час виконання розділу 3 роботи розкриваються такими положенням.

1. Чисельний експеримент проведено для перевірки розроблених алгоритмів та прогнозованих результатів їх роботи на реальних сигналах електрокардіограми пацієнта.

2. Для оцінки результативності прийняття рішень вводилися та реалізовувалися на програмному рівні статистичні критерії специфічності та чутливості .

3. Чисельний експеримент проводився на даних з мережі Internet, для роботи з якими було написано додаткову процедуру, що виконує задачу фрагментації сигналу, синхронізації QRS-комплексів у вікні спостережень, утворюючи потрібну вибірку і створюючи тим самим умови для забезпечення можливості порівняння форми сигналів під час їх розпізнавання та формування потрібних еталонів.

4. Реалізовано процедури статистичних підрахунків та візуалізації змін у кількості правильно прийнятих рішень щодо типів сигналів та в критеріях чутливості та специфічності їх розпізнавання. Візуалізація результатів супроводжується ілюстраціями стовпчиковими діаграмами.

Загальний аналіз результатів показав, що розроблений алгоритм спроможний до підвищення результативності роботи розпізнавальних систем. Прогнозовані результати справдилися. Особливо це стосується твердження про те, що узгодження між метрикою кластеризації та кореляційним коефіцієнтом в розпізнавальному алгоритмі сприяє підвищенню вірності розпізнавання типів сигналів.

Розроблена процедура призвела до зменшення кількості помилкових рішень, як для сигналів різних типів, що збігаються на окремих ділянках, так і для сигналів, яким характерний високий розкид та розбіжність зразків сигналів.

Проведені дослідження можуть застосовуватися у сфері медичної діагностики та повноцінно апробовані на реальних медико-біологічних сигналах.

РОЗРОБКА СТАРТАП-ПРОЕКТУ

4.1 Опис ідеї проекту

Для успішного впровадження проекту необхідно детально проаналізувати зміст ідеї, напрямки розповсюдження проекту та основні його переваги [20-23]. Основна ідея стартап-проекту зведена до таблиці 4.1.

Таблиця 4.1 – Опис ідеї стартап-проекту

Зміст ідеї	Напрямки застосування	Вигоди для користувача
Розробка системи розпізнавання медико-біологічних сигналів, що навчається комбінованим методом	Медична діагностика	Висока результативність роботи алгоритму у порівнянні зі стандартними способами навчання систем розпізнавання, без значного збільшення затрат

Для просування проекту необхідно оцінити його характеристики, сильні та слабкі сторони. Окрім цього необхідно провести об'єктивний аналіз конкурентних продуктів та засобів.

У наступну таблицю 4.2 зведено аналіз потенційних техніко-економічних переваг ідеї, в порівнянні з конкурентними засобами. Оцінка характеристик конкурентів наводиться чисельним значенням від 1 до 5 балів.

Отже, сильними сторонами подібних проектів є ефективність роботи та прозорість результатів роботи системи, для подальшого аналізу дослідником.

Таблиця 4.2 – Визначення сильних, слабких та нейтральних характеристик ідеї проекту

№ п/п	Техніко-економічні характеристики ідеї	(потенційні) товари/концепції конкурентів			W (слабка сторона)	N (нейтральна сторона)	S (сильна сторона)
		Мій проект	Система навчання з учителем	Система самонавчання			
1	Ефективність роботи	5	3	4			+
2	Простота застосування	4	5	4		+	
3	Прозорість результатів	5	5	1			+
4	Ресурсо-затратість	4	5	5		+	

Отже, даний підрозділ роботи розкриває ідею стартап-проекту, що полягає у розробці та впровадження системи розпізнавання медико-біологічних сигналів, що навчається комбінованим методом у сферу медичної діагностики пацієнтів.

Результатом роботи є визначення основної мети та змісту проекту, та оцінка його можливих конкурентів, з врахуванням слабких, нейтральних та сильних сторін запропонованого проекту [20-23].

4.2 Технологічний аудит ідеї проекту

Розробка стартап-проекту потребує проведення технологічного аудиту проекту, який полягає у оцінці використаних у проекті технологій, та загальній їх наявності. Результати технологічного аудиту проекту розробки системи розпізнавання медико-біологічних сигналів, що навчається комбінованим методом наведено у таблиці 4.3.

Таблиця 4.3 – Технологічна здійсненність ідеї проекту

№ п/п	Ідея проекту	Технології її реалізації	Наявність технологій	Доступність технологій
1	Комбінування алгоритмів машинного навчання	Процедура навчання з учителем	Наявна	Доступна
		Кластеризація k- means	Наявна	Доступна
2	Процедура розпізнавання	Процедура розпізнавання при навчанні з учителем	Наявна, потребує доопрацювання	Доступна
3	Програмна реалізація	Програмне середовище MatLab	Наявна	Доступна за додаткову плату
<p>Обрана технологія реалізації ідеї проекту: комбінування відомих апробованих алгоритмів із відповідним доопрацюванням розпізнавальної системи.</p> <p>Програмна реалізація у середовищі MatLab.</p>				

Проект реалізується на основі відомих алгоритмів із відповідним доопрацюванням. Алгоритм підходить для реалізації у більшості програмних середовищ, тому можна уникнути додаткової плати за програмне забезпечення.

MatLab обрано з метою простоти та наявності засобів для роботи із медико-біологічними сигналами.

4.3 Аналіз ринкових можливостей запуску стартап-проекту

Для повноцінного аналізу ринкових можливостей проекту необхідно охарактеризувати потенційну аудиторію [22, 23]. Загальна характеристика ринку наведена у таблиці 4.4.

Таблиця 4.4 – Попередня характеристика потенційного ринку стартап-проекту

№ п/п	Показники стану ринку (найменування)	Характеристика
1	Кількість головних гравців, од	1-5
2	Загальний обсяг продаж, грн/ум.од	1 млн. грн
3	Динаміка ринку (якісна оцінка	Зростає
4	Наявність обмежень для входу (вказати характер обмежень)	Відсутність аналогів за класом точності
5	Специфічні вимоги до стандартизації та сертифікації	Загальна сертифікація медичних виробів (техніки)
6	Середня норма рентабельності в галузі (або по ринку), %	120%

За попереднім оцінюванням ринок є привабливим для введення такого проекту. Показник рентабельності перевищує значення у 100%, а відсутність аналогів сприяє створенню потреб ринку. Детальніше аналіз ринкових потреб та цільової аудиторії наведено у таблиці 4.5.

Фактори, що загрожують та сприяють ринковому впровадженню проекту зведено до таблиць 4.6 – 4.7.

Таблиця 4.5 – Характеристика потенційних клієнтів стартап-проекту

№ п/п	Потреба, що формує ринок	Цільова аудиторія (цільові сегменти ринку)	Відмінності у поведінці різних потенційних цільових груп клієнтів	Вимоги споживачів до товару
1	Спосіб машинного навчання, що забезпечуватиме детальне та вірне розпізнавання медико-біологічних сигналів,	Медичні заклади, дослідники	Відсутні, проект одно-направлений	Якісне та безпомилкове розпізнавання медико-біологічних сигналів

Таблиця 4.6 – Фактори загроз

№ п/п	Фактор	Зміст загрози	Можлива реакція компанії
1	Ринок збуту	Обмежена сфера застосування проекту	Пошук альтернативних сфер застосування
2	Фінанси	Втрата фінансування та низький рівень продажу	Залучення інвесторів та просування проекту

Конкуренція не створює загрози стартап-проекту, так як інші засоби аналізу сигналів що застосовуються у передбачуваній сфері суттєво відрізняються за способом роботи та класом точності, хоч і використовують схожу методику.

Таблиця 4.7 – Фактори можливостей

№ п/п	Фактор	Зміст можливості	Можлива реакція компанії
1	Функціонал	Відсутність повноцінних аналогів.	Робота з постійними споживачами.
2	Вимогливість	Відносно мала ресурсо-затратність проекту.	Удосконалення продукту.
3	Фінанси	Зріст рівня продажу та підвищення інвестицій	Залучення партнерів, створення програм

Пропозиції та загальні риси продукту на ринку зведено до таблиці 4.8.

Таблиця 4.8 – Ступеневий аналіз конкуренції на ринку

Особливості конкурентного середовища	В чому проявляється дана характеристика	Вплив на діяльність підприємства (можливі дії компанії, щоб бути конкурентоспроможною)
1. Вказати тип конкуренції: олігополія	Конкуренція з декількома великими напрямками	Конкурентоспроможність гарантується перевагами продукту
2. За рівнем конкурентної боротьби: міжнародна	Конкуренти не обмежені країною	Адаптація продукту під ринки окремих країн
3. За галузевою ознакою: внутрішньогалузева	Сфера застосування обмежена однією галуззю	Розвиток сфери застосування системи у перспективі
4. Конкуренція за видами товарів: товарно-видова	Призначення конкурентних товарів аналогічне	Урахування переваг перед конкурентами

Продовження таблиці 4.8

5. За характером конкурентних переваг: нецінова	Основна перевага продукту – його можливості	Покращення можливостей із додаванням нових процедур
6. За інтенсивністю: марочна	Ідея представляється лише одним брендом	Покращення характеристики бренду

До таблиці 4.9 зведено характеристики основних конкурентів ідеї.

Таблиця 4.9 – Аналіз конкуренції в галузі за М. Портером

Складові аналізу	Прямі конкуренти в галузі	Потенційні конкуренти	Постачальники	Клієнти	Товари замітники
	Компанії, що розробляють процедури на основі системи навчання з учителем	Компанії, що розробляють процедури, на основі інших методів	Велика кількість	Обмежена кількість	Відсутні
Висновки:	Конкуренція інтенсивна, але передбачува на рядом переваг	Конкуренція незначна, поверхнева	Позитивно впливає на конкурентність	Сфера вживання обмежена	Проект незамінний

Отже, за результатами аналізу цільової аудиторії, ринкової ситуації та можливих конкурентів можна зробити висновок про можливість впровадження проекту на ринок, навіть за наявності конкурентів, так як ідея проекту є новою і не повторюваною до цього часу.

До наступної таблиці 4.10 зведено аналіз загальної конкурентоспроможності

Таблиця 4.10 – Обґрунтування факторів конкурентоспроможності

№ п/п	Фактор конкурентоспроможності	Обґрунтування
1	Простота впровадження	Не потребує додаткового переобладнання
2	Універсальність	Можливість застосування у парі з використанням відомих процедур
3	Ціна	Не потребує використання додаткових засобів та приладів, працює лише на програмному рівні

проекту із урахуванням попиту, споживачів та факторів маркетингового середовища. Аналіз слабких та сильних сторін проекту, та їх впливу, ілюструється у таблиці 4.11.

Таблиця 4.11 – Порівняльний аналіз сильних та слабких сторін проекту

№ п/п	Фактор конкурентоспроможності	Бали 1-20	Рейтинг товарів-конкурентів у порівнянні з проектом системи розпізнавання сигналів комбінованим методом						
			-3	-2	-1	0	1	2	3
1	Відсутність аналогів	20			+				
2	Простота впровадження	15						+	
3	Ціна	10				+			

Отже, проаналізувавши конкурентоспроможність стартапу можна зробити висновок, що імовірність реалізації та споживання проекту є дуже високою.

Фінальним етапом аналізу конкурентоспроможності є SWOT-аналіз, що допоможе підсумувати результати роботи (табл. 4.12).

Таблиця 4.12 – SWOT- аналіз стартап-проекту

Сильні сторони: Результативність роботи, ціна, універсальність, простота впровадження та застосування	Слабкі сторони: Збільшене використання розрахункової потужності у порівнянні зі стандартними методами
Можливості: Застосування продукту у різних сферах медичної діагностики. Можливе розширення сфери вживання у перспективах	Загрози: Фактично відсутні, через відсутність повноцінних аналогів.

На основі проведення аналізу можна розрізнити декілька альтернатив впровадження продукту (табл. 4.13).

Таблиця 4.13 – Альтернативи ринкового впровадження стартап-проекту

№ п/п	Альтернатива (орієнтовний комплекс заходів) ринкової поведінки	Ймовірність отримання ресурсів	Строки реалізації
1	Розробка самот. апаратної бази	75%	3-5 місяців
2	Розширення спеціалізації	60%	Термін невідомий
3	Розробка мобільного додатку	30%	1-3 місяці

Так як ідея орієнтована на обмежену сферу застосування то альтернативні способи впровадження здебільше є лише теоретичними. Єдиним повноцінним варіантом є розробка самотійного пристрою, як додатку до стандартних медичних реєстраторів сигналів. Розширення спеціалізації проекту і вихід за межі медичної діагностики може призвести до ймовірного підвищення кількості споживачів, але складність розробки та терміни здійснення достатньо великі. Розробка мобільного додатку можлива лише за виконання другого підпункту таблиці.

4.4 Аналіз ринкових можливостей запуску стартап-проекту

Найпершою процедурою з аналізу ринкових можливостей проекту є дослідження груп потенційних споживачів, що у певному сенсі визначатиме напрямки можливої подальшої реалізації проекту. Результати аналізу цільових груп зведено до таблиці 4.14.

Таблиця 4.14 – Вибір цільових груп потенційних споживачів

№ п/п	Опис профілю цільової групи потенційних клієнтів	Готовність споживачів сприйняти продукт	Орієнтовний попит в межах цільової групи (сегменту)	Інтенсивність конкуренції в сегменті	Простота входу у сегмент
1	Медичні установи	Висока	Високий	Низька	Середня
2	Дослідники та науковці	Висока	Середній	Середня	Середня

В результаті визначено стратегію охоплення ринку – концентрований маркетинг, так як сфера застосування ідеї обмежена. У далеких перспективах планується розширення сфери застосування і можлива зміна стратегії.

Для роботи в обраних сегментах ринку необхідно визначити базову стратегію розвитку стартап-проекту (табл. 4.15—4.16). Загальна стратегія визначена як концентрований маркетинг, так як продукт призначений для певної сфери діяльності. Альтернативні стратегії є аналогічними [23].

Таблиця 4.15 – Визначення базової стратегії розвитку

№ п/п	Обрана альтернатива розвитку проекту	Стратегія охоплення ринку	Ключові конкурентоспроможні позиції відповідно до обраної альтернативи	Базова стратегія розвитку*
1	Розробка самостійної апаратної бази	Концентрований маркетинг	Повне виключення необхідності застосування додаткової апаратури (окрім реєстраторів)	Стратегія спеціалізації

Таблиця 4.16 – Визначення базової стратегії конкурентної поведінки

№ п/п	Чи є проект «першопрохідцем» на ринку	Чи буде компанія шукати нових споживачів, або забирати існуючих у конкурентів?	Чи буде компанія копіювати основні характеристики товару конкурента, і які?	Стратегія конкурентної поведінки*
1	В певному сенсі є першопрохідцем	У більшості випадків – забирати існуючих	Використання загальної характеристики вірності розпізнавання із відповідним покращенням	Стратегія виклику лідера

У залежності від обраної стратегії конкурентної поведінки, та на основі вимог споживачів – визначається стратегія позиціонування [20-23] (табл 4.17).

Таблиця 4.17 – Визначення стратегії позиціонування

№ п/п	Вимоги до товару цільової аудиторії	Базова стратегія розвитку	Ключові конкурентоспроможні позиції власного стартаппроекту	Вибір асоціацій, які мають сформувати комплексну позицію власного проекту (три ключових)
1	Результативність	Стратегія спеціалізації	Точність роботи	Продуктивність
2	Вибагливість проекту		Уніфікація з іншими пристроями та базами	Простота
3	Доступність		Зниження вартості	Ціна

Отже, під час розробки підрозділу було визначено основні можливості стартап проекту на ринку, основна стратегія його розвитку та конкурентної поведінки.

Як базову стратегію розвитку було обрано стратегію спеціалізації [20, 22], так як проект розраховано на окрему цільову групу споживачів.

Стратегія конкурентної поведінки – виклик лідера.

У наступному підрозділі розкривається розробка маркетингової програми стартапу.

4.5 Розроблення маркетингової програми стартап-проекту

Найпершим кроком розроблення маркетингової програми стартапу є аналіз ключових переваг концепції товару (табл 4.18).

Таблиця 4.18 – Визначення базової стратегії конкурентної поведінки

№ п/п	Потреба	Вигода, яку пропонує товар	Ключові переваги перед конкурентами (існуючі або такі, що потрібно створити)
1	Точний інструмент для розпізнавання сигналів	Можливість розпізнавання з високою достовірністю	Комбінований алгоритм навчання розпізнавальної системи
2	Мінімальна вимогливість	Проект не потребує додаткового устаткування	Менша вартість продукту

Для подальшого маркетингового розвитку розробляється трирівнева модель товари [23], у якій підводяться підсумки щодо прогнозованих характеристик товару, та їх реального втілення. Трирівнева маркетингова модель стартап-проекту зведена до таблиці 4.19.

Відповідно до отриманої характеристики, основні переваги продукту перед аналогами – на технічному рівні. Невисока собівартість та вимогливість продукту сприяє появі економічних переваг.

Таблиця 4.19 – Опис трьох рівнів моделі товару

Рівні товару	Сутність та складові		
I. Товар за задумом	Система розпізнавання медико-біологічних сигналів		
II. Товар у реальному виконанні	Властивості/характеристики	М/Нм	Вр/Тх /Тл/Е/Ор
	1. Висока достовірність результатів розпізнавання	Нм	Тх
	2. Точність урахування особливостей розподілу точок сигналу	Нм	Т
	3. Не потребує додаткового устаткування	М	Е
	Якість: апробовано на прикладі розпізнавання QRS-комплексів реального сигналу електрокардіограми пацієнта		
	Пакування: відсутнє		
Назва товару: Система розпізнавання медико-біологічних сигналів, що навчається комбінованим методом			
III. Товар із підкріпленням	До продажу: базова версія		
	Після продажу: технічна підтримка та модернізація		
За рахунок чого потенційний товар буде захищено від копіювання: захист інтелектуальної власності			

Наступний крок – визначення цінових меж продукту [23]. Орієнтовний ціновий діапазон наведено у таблиці 4.20.

У таблиці 4.21 визначено основну систему та канали збуту товару. Продукт розповсюджується за нуль- та однорівневим каналом збуту [20], за допомогою посередників та власними силами.

Таблиця 4.20 – Визначення меж встановлення ціни

№ п/п	Рівень цін на товари- замінники	Рівень цін на товари- аналоги	Рівень доходів цільової групи споживачів	Верхня та нижня межі встановлення ціни на товар/послугу
1	5 – 15 тис. грн	10 – 30 тис. грн	5 – 25 тис. грн	3 тис. грн / 6 тис. грн

Таблиця 4.21 – Формування системи збуту

№ п/п	Специфіка закупівельної поведінки цільових клієнтів	Функції збуту, які має виконувати постачальник товару	Глибина каналу збуту	Оптимальна система збуту
1	Потреба у товарі за найвищими характеристиками	Постачання товару та зворотній зав'язок із виробником	Нуль- та однорівневий	Самостійно та із залученням посередника

Фінальним етапом маркетингової програми є складання концепції маркетингових комунікацій, відносно основної поведінки цільової аудиторії споживачів. Основні положення розробленої концепції зведено до таблиці 4.22.

Отже, на даному етапі розробка маркетингової частини стартап-проекту вважається завершеною. В результаті розробки даного підрозділу маємо чітко сформовану ринкову модель стартап-проекту, можливий її розвиток та загальну оцінку ринкового середовища, цільової групи споживачів і їх потреб.

Таблиця 4.22 – Концепція маркетингових комунікацій

№ п/п	Специфіка поведінки цільових клієнтів	Канали комунікацій, якими користуються цільові клієнти	Ключові позиції, обрані для позиціонування	Завдання рекламного повідомлення	Концепція рекламного звернення
1	Розширена професійна діяльність	Мережа Internet, професійне середовище	Моніторинг аналогів, порівняння характеристик	Увага на характеристиках	Висока якість у порівнянні з аналогами

4.6 Висновки до розділу «Розробка стартап-проекту»

Отже. Під час розробки розділу було здійснено загальну оцінку можливості утворення самостійного продукту на основі досліджень цієї роботи.

1. Проект можливо комерціалізувати, про що свідчить рентабельність проекту. За попередніми прогнозами, попит на товар з описаними характеристиками – значний.

2. З огляду на потенційні групи клієнтів, перспективи впровадження товару досить високі, оскільки товар має високу конкурентоздатність при мінімальній кількості конкурентів. Доцільна подальша імплементація проекту на рівні галузі.

3. Розробка проекту можлива як у якості ідеї, так і альтернативним варіантом, у вигляді самостійного пристрою.

ВИСНОВКИ

На виконання завдання щодо комбінованого машинного навчання процедур розпізнавання медико-біологічних сигналів проведено такі дослідження та розробки.

1. Розроблено і програмно реалізовано в середовищі MatLab спосіб комбінованого машинного навчання процедур розпізнавання медико-біологічних сигналів за їх формою в межах заданої медично інтерпретованої класифікації з комбінуванням навчання з учителем із самонавчанням шляхом кластеризації навчальних вибірок цих сигналів.

2. Програмно реалізовано розпізнавальний алгоритм для перевірки результативності його навчання запропонованим способом відповідно до заданих типів медико-біологічних сигналів, кількості типів в їх класифікації, формату даних навчальних вибірок та складу ознак для розпізнавання сигналів у вигляді послідовностей відліків характеристики їх форми на рівномірній сітці на шкалі часу.

3. Забезпечена синхронізація вибірки сигналів із запису даних пацієнта в Інтернет в єдиному часовому вікні з рівномірною шкалою відліків для їх спостереження, а також для формування навчальних і контрольних вибірок, порівняння форми сигналів в цілях визначення їх типів, розрахунку їх еталонних зразків для досліджуваних класів і для сформованих відповідно до способу навчання кластерів. Можливість такої синхронізації і використання запису сигналів для формування вибірок, що потрібні для проведення досліджень, забезпечена достатньою тривалістю запису сигналів в базі даних в Інтернет і відповідною розміткою цього запису.

4. З використанням відповідних програмних процедур проведено формування навчальних вибірок для реалізації навчання обраного для досліджень розпізнавального алгоритму розробленим комбінованим способом із

використанням багато еталонного опису класів сигналів, що розпізнаються. Цей опис формується навчальним алгоритмом відповідно до апріорних даних про особливості ймовірнісного розподілу сигналів у вибірках.

5. На тестовому прикладі оброки запису сигналів пацієнта статистичним способом підрахунку правильних та помилкових рішень на контрольній вибірці перевірена результативність навчання розпізнавального алгоритму запропонованим комбінованим способом. Виконана порівняльна оцінка показників чутливості і специфічності розпізнавання сигналів до і після навчання з використанням двох різних метрик кластеризації навчальних вибірок.

За результатами проведених досліджень та розробок спроможними можливо вважати такі висновки.

1. Попередня фільтрація сигналів навчальних і контрольних вибірок для стабілізації ізолінії, на якій розташовуються QRS-комплекси трьох типів, які розглядаються як сигнали, що розпізнаються, а також перехід від представлення таких сигналів у вигляді послідовностей їх відліків в запису до послідовностей відліків характеристик їх форми, як показали попередні дослідження, сприяють покращенню умов розпізнавання сигналів в тестовому прикладі.

2. Більш детальне врахування особливостей ймовірнісного розподілу сигналів, що розпізнаються за їх формою, в процесі машинного комбінованого навчання процедур розпізнавання медико-біологічних сигналів запропонованим способом, як показав тестовий приклад обробки даних пацієнта, може сприяти підвищенню чутливості і специфічності визначення типів сигналів заданої класифікації,

3. Кластеризація навчальних вибірок, яка передбачена розробленими процедурами комбінованого навчання, забезпечує отримання адекватного багато еталонного опису цих вибірок згідно з реальними особливостями ймовірнісних розподілів сигналів в апріорних даних цих вибірок, про що свідчать результати статистичних досліджень в тестовому прикладі на контрольних вибірках.

4. Якість комбінованого навчання, яке розглянуто в роботі, валідність рішень щодо визначення типів сигналів заданої класифікації за їх формою розробленими

програмними процедурами, за показниками чутливості і специфічності помітно, на декілька відсотків, залежить від метрик, що використовуються в алгоритмах кластеризації навчальних вибірок.

5. Узгодження метрик в процедурах кластеризації вибірок сигналів, які є основою алгоритмів навчання розпізнавальних процедур, з метрикою, що використовується в критерії визначення типів сигналів, які розпізнаються (скалярним добутком відповідних векторів), може призвести до підвищення чутливості і специфічності прийняття рішень, на що вказують результати статистичних досліджень на контрольних вибірках.

6. В цілому, чисельні статистичні дослідження результативності комбінованого навчання процедур розпізнавання медико-біологічних сигналів за їх формою на тестовому прикладі підтвердили спроможність виконаних в роботі досліджень і розробок.

7. В цілому, надані алгоритми і програмні процедури навчання для алгоритмів розпізнавання сигналів за їх формою, які легко конкретизувати і переналагодити на іншу кількість типів сигналів в їх класифікації, на їх конкретний вигляд і формат представлення, доцільно розглядати як додатковий інструментарій для проведення досліджень щодо виявлення наявності, оцінки і реалізації можливих резервів у підвищенні валідності розпізнавання сигналів за їх формою в системах аналізу даних пацієнтів в цілях їх діагностики.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Васильев В.И. Распознающие системы. Справочник. Киев, Наукова думка, 1983. – 422 с.
2. Генкин А.А. Новая информационная технология анализа медицинских данных (программный комплекс ОМИС) / А.А. Генкин – СПб.: Политехника, 1999. - 191 с.
3. Васильев В.И. Распознающие системы. Справочник. Киев, Наукова думка, 1983. – 422 с.
4. Estivill-Castro, Vladimir (20 June 2002). Why so many clustering algorithms – A Position Paper. ACM SIGKDD Explorations Newsletter 4 (1): 65–75. doi:10.1145/568574.568575.
5. Дюк В. Информационные технологии в медико-биологических исследованиях./ В. Дюк, В. Эммануэль. - СПб.: Питер, 2003. –528 с. – ISBN I978-5-94723-501-2
6. Коростелев А.А., Ключев Н.Ф., Мельник Ю.А. и др. Теоретические основы радиолокации: Учебное пособие для вузов. Под ред. В. Е. Дулевича. - 2-е изд., перераб. и доп. - М., Сов. радио, 1978. - 608 с.
7. Розенблатт, Ф. Принципы нейродинамики: Перцептроны и теория механизмов мозга = Principles of Neurodynamic: Perceptrons and the Theory of Brain Mechanisms. — М.: Мир, 1965. — 480 с.
8. Лагутін В. В., Сиротенко В. І., Шачиков А. Д., Шуляк О. П. , " Кластеризація медико-біологічних сигналів в розпізнавальних системах, що навчаються з учителем" «Мікросистеми, Електроніка та Акустика» , Київ, Україна, 2019, 38 - 52 с, DOI: 10.20535/2523-4455.2019.24.6
9. A. Shulyak, T. Saurova, A. Shachykov, V. Sirotenko and V. Lahutin, "Modifications of recognition procedures of biomedical signals using shape

- characteristics for their description and in decision criteria," 2020 IEEE 40th International Conference on Electronics and Nanotechnology (ELNANO), Kyiv, Ukraine, 2020, pp. 520-525, DOI: 10.1109/ELNANO50318.2020.9088862
10. Лагутін В. В., Сиротенко В. І., Шачиков А. Д., Шуляк О. П. , "Вибір меж для шкал медико-біологічних сигналів в алгоритмах їх розпізнавання " «Мікросистеми, Електроніка та Акустика» , Київ, Україна, 2020, 11 - 19 с, DOI: 10.20535/2523-4455.me.197291
 11. A.P. Shulyak, A.D.Shachykov. (2016) “Criteria and Procedures for Estimating the Informativity and Feature Selection in Biomedical Signals for their Recognition”. «Вісник НТУУ КПІ». Серія: Радіотехніка. Радіоапаратобудування. № 66, с. 79-86. URL: <https://ela.kpi.ua/handle/123456789/22801>
 12. A. D. Shachykov, A. P. Shulyak. «About the Impact of Informative Features Selection in the Mutually Orthogonal Decompositions of Biomedical Signals for their Recognition». 2016 IEEE 36th International Conference Electronics and Nanotechnology (ELNANO), Kyiv, Ukraine, April 19-21, 2016, p. 228-231. doi: 10.1109/ELNANO.2016.7493054
 13. Шуляк А.П., Шачиков А.Д. Особливості використання характеристики форми медико-біологічних сигналів при їх розпізнаванні/ Шачиков А.Д., Шуляк А.П., Збірник наукових праць «ВІСНИК НТУУ «КПІ». Серія приладобудування», випуск 51(1), К.: 2016 р. с. 103-109. [https://doi.org/10.20535/1970.51\(1\).2016.78217](https://doi.org/10.20535/1970.51(1).2016.78217).
 14. А. Д. Шачиков, В. А. Шелофаст, А. П. Шуляк, “Модификации процедур отбора признаков циклических медико-биологических сигналов для их распознавания”, Вісник Національного технічного університету України Київський політехнічний інститут. Серія: Приладобудування 57 (1), К.: 2019 р. с. 103–109. DOI: 10.20535/1979.2017.53.1.016
 15. I. H. Witten, E. Frank Data Mining: Practical Machine Learning Tools and Techniques (Second Edition). — Morgan Kaufmann, 2005

16. PhysioNet. St.-Petersburg Institute of Cardiological Technics 12-lead Arrhythmia Database. URL: <http://www.physionet.org/cgi-bin/atm/ATM>
17. Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* 101(23):e215-e220 [Circulation Electronic Pages; <http://circ.ahajournals.org/cgi/content/full/101/23/e215>]; 2000 (June 13).
18. Потемкин В.Г. Введение в MATLAB. – М.: Диалог-МИФИ, 2000
19. Дьяконов В. П. MATLAB. Полный самоучитель. – М.: ДМК Пресс, 2012. – 768 с.
20. Тиль, П. От нуля к единице : как создать стартап, который изменит будущее / П. Тиль, Б. Мастерс; перевод с англ. – Москва : Альпина паблишер, 2015. – 188 с.
21. Харниш, В. Правила прибыльных стартапов : как расти и зарабатывать деньги / В. Харниш ; пер. с англ. В. Хозинского. – Москва : Манн, Иванов и Фербер, 2012. – 279 с.
22. Экланд С. Ангелы, драконы и стервятники : как привлечь правильных инвесторов в свой стартап и сохранить бизнес / С. Экланд ; пер. с англ. О. Терентьевой. – Москва : Манн, Иванов и Фербер, 2011. – 275 с
23. Розроблення стартап-проекту [Електронний ресурс] : Методичні рекомендації до виконання розділу магістерських дисертацій для студентів інженерних спеціальностей / За заг. ред. О.А. Гавриша. – Київ : НТУУ «КПІ», 2016. – 28 с.
24. Шуляк О. П., Лагутін В. В. Векторні показники подібності форми медико-біологічних сигналів в процедурах їх розпізнавання. XIX Міжнародна науково-технічна конференція «Приладобудування: стан і перспективи», 13-14 травня 2020 р., Київ, Україна : збірник матеріалів конференції. – Київ : КПІ ім. Ігоря Сікорського, 2020. – С. 94-95.
25. Шуляк О. П., Лагутін В. В. Інтегральна характеристика форми медико-біологічних сигналів в процедурах їх розпізнавання. XIX Міжнародна

- науково-технічна конференція «Приладобудування: стан і перспективи», 13-14 травня 2020 р., Київ, Україна : збірник матеріалів конференції. – Київ : КПІ ім. Ігоря Сікорського, 2020. – С. 91-92.
26. О. П. Шуляк, В. В. Лагутін. Комбіноване навчання алгоритма розпізнавання медико-біологічних сигналів на прикладі визначення типів QRS-комплексів електрокардіограми / Актуальні проблеми автоматики та приладобудування : матеріали 2-ї Міжнар. наук.-техн. конф., 06-07 грудня 2018 р. – Харків : Панов А. М., 2018. – С. 113-114.
 27. A.K. Jain and R.C. Dubes. Algorithms for Clustering Data. Prentice Hall, 1988..
 28. Воронцов К.В. Алгоритмы кластеризации и многомерного шкалирования. Курс лекций. МГУ, 2007.
 29. Котов А., Красильников Н. Кластеризация данных. Курс лекций. МГУ, 2006.
 30. Жамбю М. Иерархический кластер-анализ и соответствия. — М.: Финансы и статистика, 1988. — 345 с.
 31. Estivill-Castro, Vladimir (20 June 2002). Why so many clustering algorithms – A Position Paper. ACM SIGKDD Explorations Newsletter 4 (1): 65–75. doi:10.1145/568574.568575.
 32. Бериков В. С., Лбов Г. С. Современные тенденции в кластерном анализе // Всероссийский конкурсный отбор обзорно-аналитических статей по приоритетному направлению «Информационно-телекоммуникационные системы», 2008. — 26 с.
 33. Т. И. Мурашкина «Техника физического эксперимента и метрология» Санкт-Петербург : Политехника, 2015. - 137, [1] с. : ил. ; 8,75 усл. печ. л. -- ISBN 978-5-7325-1051-5

Додаток А

Лістинг робочої програми попередньої обробки сигналів та формування вибірок
QRS-комплексів з неперервного запису ЕКГ пацієнта

```

clc
clear all
close all
warning('off')
set(0, 'DefaultAxesFontSize', 14, 'DefaultAxesFontName', 'Times New Roman');

%Зчитування файлів
fid = fopen ('I22m.mat');
fid1= fopen ('anot_22.txt');
ECG = fread(fid, 462600, 'int16')./623;

A=textread('anot_22.txt', '%s');
num_r=10:6:18764;
num_type=11:6:18764;

qrs_pos=zeros(size(num_r));

for i=1:length(num_r)
    qrs_pos(i)=str2num(A{num_r(i),:});
    qrs_type(i,:)=A{num_type(i),:};
end

P = 1;
Q = 40;
R = 48;
S = 52;
T = 128;

window_l = R-1;
window_r = T-R;

Na=0;
Nv=0;
Nn=0;

for i=1:length(num_r)
    if strcmp(qrs_type(i,:), 'N')
        [val, max_i]=max(ECG((qrs_pos(i)-window_l/4):(qrs_pos(i)+window_r/4)));
        max_i=qrs_pos(i)-window_l/4+max_i;
        Nn = Nn + 1;
        qrs_n(Nn, :)=ECG((max_i-window_l):(max_i+window_r));
    else if strcmp(qrs_type(i,:), 'V')
        [val, max_i]=max(ECG((qrs_pos(i)-window_l/4):(qrs_pos(i)+window_r/4)));
        max_i=qrs_pos(i)-window_l/4+max_i;
        Nv = Nv + 1;
        qrs_v(Nv, :)=ECG((max_i-window_l):(max_i+window_r));
    else if strcmp(qrs_type(i,:), 'A')
        [val, max_i]=max(ECG((qrs_pos(i)-
window_l/4):(qrs_pos(i)+window_r/4)));
        max_i=qrs_pos(i)-window_l/4+max_i;
        Na = Na + 1;
    end
end

```

```

        qrs_a(Na,:)=ECG((max_i-window_l):(max_i+window_r));
    end
end
end
end

% рівномірно обираємо потрібну кількість комплексів
% в даному випадку - половина від мінімальної кількості по усім типам
minN = min([Na,Nn,Nv])/2;
step_a = floor(Na/minN)
step_n = floor(Nn/minN)
step_v = floor(Nv/minN)

nomera = [step_a*(1:minN);step_n*(1:minN);step_v*(1:minN)]'
nomera_control = [floor(step_a*(1:minN)-step_a/2); floor(step_n*(1:minN)+4);
floor(step_v*(1:minN)-step_v/2)]' %+22 в текст

qrs_a_control=qrs_a(nomera_control(:,1),:);
qrs_n_control=qrs_n(nomera_control(:,2),:);
qrs_v_control=qrs_v(nomera_control(:,3),:);

qrs_a=qrs_a(nomera(:,1),:);
qrs_n=qrs_n(nomera(:,2),:);
qrs_v=qrs_v(nomera(:,3),:);

%-----

%обробка вибірок
for i = 1:minN
    %фільтр
    qrs_a_filtered(i,:) = medfilt1(qrs_a(i,:));
    qrs_n_filtered(i,:) = medfilt1(qrs_n(i,:));
    qrs_v_filtered(i,:) = medfilt1(qrs_v(i,:));

    qrs_a_form(i,:)= qrs_a_filtered(i,:) -
sum(qrs_a_filtered(i,:))./(window_l+window_r);
    qrs_n_form(i,:)= qrs_n_filtered(i,:) -
sum(qrs_n_filtered(i,:))./(window_l+window_r);
    qrs_v_form(i,:)= qrs_v_filtered(i,:) -
sum(qrs_v_filtered(i,:))./(window_l+window_r);
    qrs_a_norm(i,:) = qrs_a_form(i,+)/sqrt(sum(qrs_a_form(i,:).*qrs_a_form(i,:)));
    qrs_n_norm(i,:) = qrs_n_form(i,+)/sqrt(sum(qrs_n_form(i,:).*qrs_n_form(i,:)));
    qrs_v_norm(i,:) = qrs_v_form(i,+)/sqrt(sum(qrs_v_form(i,:).*qrs_v_form(i,:)));

    qrs_a_control_form(i,:)= qrs_a_control(i,:) -
sum(qrs_a_control(i,:))./(window_l+window_r);
    qrs_n_control_form(i,:)= qrs_n_control(i,:) -
sum(qrs_n_control(i,:))./(window_l+window_r);
    qrs_v_control_form(i,:)= qrs_v_control(i,:) -
sum(qrs_v_control(i,:))./(window_l+window_r);
    qrs_a_control_norm(i,:) =
qrs_a_control_form(i,+)/sqrt(sum(qrs_a_control_form(i,:).*qrs_a_control_form(i,:))
);
    qrs_n_control_norm(i,:) =
qrs_n_control_form(i,+)/sqrt(sum(qrs_n_control_form(i,:).*qrs_n_control_form(i,:))
);
    qrs_v_control_norm(i,:) =
qrs_v_control_form(i,+)/sqrt(sum(qrs_v_control_form(i,:).*qrs_v_control_form(i,:))
);

%ілюстрація обробки
figure(1)
subplot(2,1,1)

```

```

    plot(qrs_n(i,:))
    hold on
    grid on
    xlabel('N, counts')
    ylabel('U, mV')
    xlim([1,window_l+window_r])
    subplot(2,1,2)
    plot(qrs_n_norm(i,:))
    hold on
    grid on
    xlabel('N, counts')
    xlim([1,window_l+window_r])
end

%формування еталонів
etalon_a_no_form = mean(qrs_a_norm);
etalon_n_no_form = mean(qrs_n_norm);
etalon_v_no_form = mean(qrs_v_norm);

save('ECG_types/etalon_A_no_form.mat','etalon_a_no_form','-mat')
save('ECG_types/etalon_N_no_form.mat','etalon_n_no_form','-mat')
save('ECG_types/etalon_V_no_form.mat','etalon_v_no_form','-mat')

etalon_a_form = mean(qrs_a_norm) - sum(mean(qrs_a_norm)/(window_l+window_r));
etalon_a = etalon_a_form/sqrt(sum(etalon_a_form.*etalon_a_form));
etalon_n_form = mean(qrs_n_norm) - sum(mean(qrs_n_norm)/(window_l+window_r));
etalon_n = etalon_n_form/sqrt(sum(etalon_n_form.*etalon_n_form));
etalon_v_form = mean(qrs_v_norm) - sum(mean(qrs_v_norm)/(window_l+window_r));
etalon_v = etalon_v_form/sqrt(sum(etalon_v_form.*etalon_v_form));

%в файли
save('ECG_types/A_type.mat','qrs_a','-mat')
save('ECG_types/N_type.mat','qrs_n','-mat')
save('ECG_types/V_type.mat','qrs_v','-mat')

save('ECG_types/A_type_form.mat','qrs_a_norm','-mat')
save('ECG_types/N_type_form.mat','qrs_n_norm','-mat')
save('ECG_types/V_type_form.mat','qrs_v_norm','-mat')

save('ECG_types/A_type_control_form.mat','qrs_a_control_norm','-mat')
save('ECG_types/N_type_control_form.mat','qrs_n_control_norm','-mat')
save('ECG_types/V_type_control_form.mat','qrs_v_control_norm','-mat')

save('ECG_types/etalon_A_type.mat','etalon_a','-mat')
save('ECG_types/etalon_N_type.mat','etalon_n','-mat')
save('ECG_types/etalon_V_type.mat','etalon_v','-mat')

%малюємо еталони
figure(2)
subplot(3,1,1)
plot(etalon_n,'k','linewidth',2)
grid on
xlabel('N, counts')
xlim([1,window_l+window_r])

subplot(3,1,2)
plot(etalon_a,'k','linewidth',2)
grid on
xlabel('N, counts')
xlim([1,window_l+window_r])

subplot(3,1,3)
plot(etalon_v,'k','linewidth',2)

```

```

grid on
xlabel('N, counts')
xlim([1,window_l+window_r])

%приклад комплексів 3 розділ
figure(3)
plot(qrs_a_norm(8,:), 'k', 'linewidth', 2)
hold on
plot(qrs_n_norm(8,:), 'k--', 'linewidth', 2)
hold on
plot(qrs_v_norm(8,:), 'k:', 'linewidth', 2)
hold on
grid on
xlabel('N, counts')
xlim([1,window_l+window_r])

```

Додаток Б

Лістинг робочої програми алгоритму прийняття рішень про тип вхідного сигналу
у випадку навчання з учителем

```

clc
clear all
close all
warning('off')

P = 1;
Q = 40;
R = 48;
S = 52;
T = 128;

window_l = R-1;
window_r = T-R;

%підгрузка еталонів
load('ECG_types/etalon_A_type.mat','etalon_a','-mat')
load('ECG_types/etalon_N_type.mat','etalon_n','-mat')
load('ECG_types/etalon_V_type.mat','etalon_v','-mat')

%підгрузка зразків
load('ECG_types/A_type_form.mat','qrs_a_norm','-mat');
load('ECG_types/N_type_form.mat','qrs_n_norm','-mat');
load('ECG_types/V_type_form.mat','qrs_v_norm','-mat');
load('ECG_types/A_type_control_form.mat','qrs_a_control_norm','-mat');
load('ECG_types/N_type_control_form.mat','qrs_n_control_norm','-mat');
load('ECG_types/V_type_control_form.mat','qrs_v_control_norm','-mat');

P = 1;
Q = 40;
R = 48;
S = 52;
T = 128;

window_l = R-1;
window_r = T-R;
%порівнюємо за скалярним добутком
true_A = 0;
mistake_A_to_N = 0;
mistake_A_to_V = 0;
true_N = 0;
mistake_N_to_A = 0;
mistake_N_to_V = 0;
true_V = 0;
mistake_V_to_N = 0;
mistake_V_to_A = 0;
for i =
1:min([size(qrs_a_control_norm,1),size(qrs_n_control_norm,1),size(qrs_v_control_no
rm,1)])
    similarity_a(i,:) =
[dot(qrs_a_control_norm(i,:),etalon_a),dot(qrs_a_control_norm(i,:),etalon_n),dot(q
rs_a_control_norm(i,:),etalon_v)]/2 + 1/2;

```

```

        similarity_n(i,:) =
[dot(qrs_n_control_norm(i,:),etalon_a),dot(qrs_n_control_norm(i,:),etalon_n),dot(q
rs_n_control_norm(i,:),etalon_v)]/2 + 1/2;
        similarity_v(i,:) =
[dot(qrs_v_control_norm(i,:),etalon_a),dot(qrs_v_control_norm(i,:),etalon_n),dot(q
rs_v_control_norm(i,:),etalon_v)]/2 + 1/2;
        %A-тип
        if max(similarity_a(i,:)) == similarity_a(i,1)
            true_A = true_A+1;
        else if max(similarity_a(i,:)) == similarity_a(i,2)
            mistake_A_to_N = mistake_A_to_N+1;
        else if max(similarity_a(i,:)) == similarity_a(i,3)
            mistake_A_to_V = mistake_A_to_V+1;
        end
    end
end
%
%N-тип
if max(similarity_n(i,:)) == similarity_n(i,1)
    mistake_N_to_A = mistake_N_to_A+1;
else if max(similarity_n(i,:)) == similarity_n(i,2)
    true_N = true_N+1;
else if max(similarity_n(i,:)) == similarity_n(i,3)
    mistake_N_to_V = mistake_N_to_V+1;
end
end
end
%
%V-тип
if max(similarity_v(i,:)) == similarity_v(i,1)
    mistake_V_to_A = mistake_V_to_A+1;
else if max(similarity_v(i,:)) == similarity_v(i,2)
    mistake_V_to_N = mistake_V_to_N+1;
else if max(similarity_v(i,:)) == similarity_v(i,3)
    true_V = true_V+1;
end
end
end

true_A
mistake_A_to_N
mistake_A_to_V

true_N
mistake_N_to_A
mistake_N_to_V

true_V
mistake_V_to_N
mistake_V_to_A
chustv_V = (true_V*100)/size(qrs_v_control_norm,1)
spec_V = (true_V*100)/(mistake_V_to_A+mistake_V_to_N+true_V)
chustv_A = (true_A*100)/(mistake_A_to_N+mistake_A_to_V+true_A)
spec_A = (true_A*100)/(mistake_N_to_A+mistake_V_to_A+true_A)
chustv_N = (true_N*100)/(mistake_N_to_A+mistake_N_to_V+true_N)
spec_N = (true_N*100)/(mistake_A_to_N+mistake_V_to_N+true_N)

chustv_V = (true_V*100)/(mistake_V_to_A+mistake_V_to_N+true_V)
spec_V = (true_V*100)/(mistake_A_to_V+mistake_N_to_V+true_V)
true_all = true_A+true_N+true_V
chustv_all_percent =
(true_all*100)/(size(qrs_a_control_norm,1)+size(qrs_v_control_norm,1)+size(qrs_n_c
ontrol_norm,1))

```

Додаток В

Лістинг робочої програми кластеризації навчальних вибірок, утворення еталонів та процедури прийняття рішень у разі комбінованого навчання алгоритму розпізнавання

```

clc
clear all
close all
set(0,'DefaultAxesFontSize',14,'DefaultAxesFontName','Times New Roman');

%підгрузка сигналів
load('ECG_types/A_type_form.mat','qrs_a_norm','-mat')
load('ECG_types/N_type_form.mat','qrs_n_norm','-mat')
load('ECG_types/V_type_form.mat','qrs_v_norm','-mat')
load('ECG_types/A_type_control_form.mat','qrs_a_control_norm','-mat')
load('ECG_types/N_type_control_form.mat','qrs_n_control_norm','-mat')
load('ECG_types/V_type_control_form.mat','qrs_v_control_norm','-mat')

P = 1;
Q = 40;
R = 48;
S = 52;
T = 128;

window_l = R-1;
window_r = T-R;

%for clusters = 2:10
clusters = 6;

%кластеризуємо навч вибірку
[IDX,C] = kmeans(qrs_a_norm,clusters);%,'distance','correlation')
clusters_a = C;
IDX_a = IDX;

[IDX,C] = kmeans(qrs_n_norm,clusters);%,'distance','correlation')
clusters_n = C;
IDX_n = IDX;

[IDX,C] = kmeans(qrs_v_norm,clusters);%,'distance','correlation')
clusters_v = C;
IDX_v = IDX;

true_A = 0;
mistake_A_to_N = 0;
mistake_A_to_V = 0;
true_N = 0;
mistake_N_to_A = 0;
mistake_N_to_V = 0;
true_V = 0;
mistake_V_to_N = 0;
mistake_V_to_A = 0;

```

```

%xf кластерів
for j = 1:clusters
    clusters_a_form(j,:) = clusters_a(j,:) -
sum(clusters_a(j,:))./(window_l+window_r);
    clusters_n_form(j,:) = clusters_n(j,:) -
sum(clusters_n(j,:))./(window_l+window_r);
    clusters_v_form(j,:) = clusters_v(j,:) -
sum(clusters_v(j,:))./(window_l+window_r);
    clusters_a_norm(j,:) =
clusters_a_form(j,)/sqrt(sum(clusters_a_form(j,:).*clusters_a_form(j,:)));
    clusters_n_norm(j,:) =
clusters_n_form(j,)/sqrt(sum(clusters_n_form(j,:).*clusters_n_form(j,:)));
    clusters_v_norm(j,:) =
clusters_v_form(j,)/sqrt(sum(clusters_v_form(j,:).*clusters_v_form(j,:)));
end

save('ECG_types/clusters_A_norm.mat','clusters_a_norm','-mat')
save('ECG_types/clusters_N_norm.mat','clusters_n_norm','-mat')
save('ECG_types/clusters_V_norm.mat','clusters_v_norm','-mat')

%перевірка
true_A = 0;
mistake_A_to_N = 0;
mistake_A_to_V = 0;
true_N = 0;
mistake_N_to_A = 0;
mistake_N_to_V = 0;
true_V = 0;
mistake_V_to_N = 0;
mistake_V_to_A = 0;

for i =
1:min([size(qrs_a_control_norm,1),size(qrs_n_control_norm,1),size(qrs_v_control_norm,1)])
    for j = 1:clusters
        similarity_a_a(i,j) =
[dot(qrs_a_control_norm(i,:),clusters_a_norm(j,:))]/2 + 1/2;
        similarity_a_a_m(i,:) = max(similarity_a_a(i,:));
        similarity_a_n(i,j) =
[dot(qrs_a_control_norm(i,:),clusters_n_norm(j,:))]/2 + 1/2;
        similarity_a_n_m(i,:) = max(similarity_a_n(i,:));
        similarity_a_v(i,j) =
[dot(qrs_a_control_norm(i,:),clusters_v_norm(j,:))]/2 + 1/2;
        similarity_a_v_m(i,:) = max(similarity_a_v(i,:));
        similarity_a =
max(max(similarity_a_a_m,similarity_a_n_m),max(similarity_a_n_m,similarity_a_v_m))
;

        similarity_n_a(i,j) =
[dot(qrs_n_control_norm(i,:),clusters_a_norm(j,:))]/2 + 1/2;
        similarity_n_a_m(i,:) = max(similarity_n_a(i,:));
        similarity_n_n(i,j) =
[dot(qrs_n_control_norm(i,:),clusters_n_norm(j,:))]/2 + 1/2;
        similarity_n_n_m(i,:) = max(similarity_n_n(i,:));
        similarity_n_v(i,j) =
[dot(qrs_n_control_norm(i,:),clusters_v_norm(j,:))]/2 + 1/2;
        similarity_n_v_m(i,:) = max(similarity_n_v(i,:));
        similarity_n =
max(max(similarity_n_a_m,similarity_n_n_m),max(similarity_n_n_m,similarity_n_v_m))
;

        similarity_v_a(i,j) =
[dot(qrs_v_control_norm(i,:),clusters_a_norm(j,:))]/2 + 1/2;

```



```

        similarity_v_a_m(i,:) = max(similarity_v_a(i,:));
        similarity_v_n(i,j) =
[dot(qrs_v_control_norm(i,:),clusters_n_norm(j,:))]/2 + 1/2;
        similarity_v_n_m(i,:) = max(similarity_v_n(i,:));
        similarity_v_v(i,j) =
[dot(qrs_v_control_norm(i,:),clusters_v_norm(j,:))]/2 + 1/2;
        similarity_v_v_m(i,:) = max(similarity_v_v(i,:));
        similarity_v =
max(max(similarity_v_a_m,similarity_v_n_m),max(similarity_v_n_m,similarity_v_v_m))
;

    end
    if similarity_a(i,:) == similarity_a_a_m(i,:)
    true_A = true_A+1;
    else if similarity_a(i,:) == similarity_a_n_m(i,:)
        mistake_A_to_N = mistake_A_to_N+1;
    else if similarity_a(i,:) == similarity_a_v_m(i,:)
        mistake_A_to_V = mistake_A_to_V+1;
    end
    end
end
if similarity_n(i,:) == similarity_n_a_m(i,:)
    mistake_N_to_A = mistake_N_to_A+1;
else if similarity_n(i,:) == similarity_n_n_m(i,:)
    true_N = true_N+1;
    else if similarity_n(i,:) == similarity_n_v_m(i,:)
        mistake_N_to_V = mistake_N_to_V+1;
    end
    end
end
if similarity_v(i,:) == similarity_v_a_m(i,:)
    mistake_V_to_A = mistake_V_to_A+1;
else if similarity_v(i,:) == similarity_v_n_m(i,:)
    mistake_V_to_N = mistake_V_to_N+1;
    else if similarity_v(i,:) == similarity_v_v_m(i,:)
        true_V = true_V+1;
    end
    end
end
end

true_A
mistake_A_to_N
mistake_A_to_V

true_N
mistake_N_to_A
mistake_N_to_V

true_V
mistake_V_to_N
mistake_V_to_A

chustv_V = (true_V*100)/size(qrs_v_control_norm,1)
spec_V = (true_V*100)/(mistake_V_to_A+mistake_V_to_N+true_V)

chustv_A = (true_A*100)/(mistake_A_to_N+mistake_A_to_V+true_A)
spec_A = (true_A*100)/(mistake_N_to_A+mistake_V_to_A+true_A)

chustv_N = (true_N*100)/(mistake_N_to_A+mistake_N_to_V+true_N)
spec_N = (true_N*100)/(mistake_A_to_N+mistake_V_to_N+true_N)

chustv_V = (true_V*100)/(mistake_V_to_A+mistake_V_to_N+true_V)
spec_V = (true_V*100)/(mistake_A_to_V+mistake_N_to_V+true_V)

```

```

true_all = true_A+true_N+true_V
chustv_all_percent =
(true_all*100)/(size(qrs_a_control_norm,1)+size(qrs_v_control_norm,1)+size(qrs_n_c
ontrol_norm,1))
%ілюстрація для 2 розділу
load('ECG_types/etalon_V_type.mat','etalon_v','-mat')
figure(1)
for i = 1:size(qrs_v_control_norm,1)
    for j = 1:clusters
        plot(qrs_v_control_norm(i,:))
        hold on
        grid on
        plot(etalon_v,'k','linewidth',3)
        hold on
        plot(clusters_v_norm(j,:), 'k--', 'linewidth',2)
        hold on
        xlabel('N, counts')
        xlim([1,window_l+window_r])
    end
end
%розподіл елементів по кластерам
figure(2)
subplot(3,1,1)
bar(IDX_a)
grid on
xlabel('Number of sample')
ylabel('Number of clusters')
subplot(3,1,2)
bar(IDX_n)
grid on
xlabel('Number of sample')
ylabel('Number of clusters')
subplot(3,1,3)
bar(IDX_v)
grid on
xlabel('Number of sample')
ylabel('Number of clusters')
vklasterah_A = hist(IDX_a)
vklasterah_N = hist(IDX_n)
vklasterah_V = hist(IDX_v)
%МАЛЮЄМО ЕТАЛОНИ
for i = 1:clusters
    figure(3)
    subplot(3,1,1)
    plot(clusters_a_norm(i,:), 'k')
    grid on
    hold on
    xlabel('N, counts')
    xlim([1,window_l+window_r])
    subplot(3,1,2)
    plot(clusters_n_norm(i,:), 'k')
    grid on
    hold on
    xlabel('N, counts')
    xlim([1,window_l+window_r])
    subplot(3,1,3)
    plot(clusters_v_norm(i,:), 'k')
    grid on
    hold on
    xlabel('N, counts')
    xlim([1,window_l+window_r])
end

```